



Modeling and imaging of the vocal fold vibration for voice health.

Granados, Alba

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Granados, A. (2016). *Modeling and imaging of the vocal fold vibration for voice health*. Technical University of Denmark, Department of Electrical Engineering.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Alba Granados Corsellas

Modeling and imaging of the vocal fold vibration for voice health

PhD thesis, September 2016

Modeling and imaging of the vocal fold vibration for voice health

Ph.D. thesis by
Alba Granados Corsellas

Technical University of Denmark
2016

This thesis was submitted to the Technical University of Denmark (DTU) as partial fulfillment of the requirements for the degree of Doctor of Philosophy (Ph.D.) in Electronics and Communication. The work presented in this thesis was completed between October 15, 2012 and July 7, 2016 at Acoustic Technology, Department of Electrical Engineering, DTU, under the supervision of Associate Professor Jonas Bruns kog. The project was partly financed by the Swedish organization AFA Försäkring (2/3) and by the Technical University of Denmark (1/3).

Title

Modeling and imaging of the vocal fold vibration for voice health

Author

Alba Granados Corsellas

Supervisor

Assoc. Prof. Jonas Bruns kog

Acoustic Technology
Department of Electrical Engineering
Technical University of Denmark
Kongens Lyngby, Denmark

Abstract

Identification of abnormalities on the vocal fold by means of different diagnostic methods is a key step to determine the cause or causes of a voice disorder, and subsequently give an adequate treatment. To this end, clinical investigations benefit from accurate mathematical models for prediction, analysis and inference. This thesis deals with biomechanical models of the vocal fold, specially of the collision, and laryngeal videoendoscopic analysis procedures suitable for the inference of the underlying vocal fold characteristics.

The first part of this research is devoted to frictionless contact modeling during asymmetric vocal fold vibration. The prediction problem is numerically addressed with a self-sustained three-dimensional finite element model of the vocal fold with position-based contact constraints. A novel contact detection mechanism is shown to successfully detect collision in asymmetric oscillations. Optimization procedures for constraint enforcement are studied to improve the accuracy of the predictions as an alternative to classical spring-like contact forces. The second part of this research investigates a non-invasive procedure to quantitatively analyze the two-dimensional vocal fold displacements captured with laryngeal high-speed videoendoscopy. A dense optical flow algorithm is adapted to the complex nature of the image sequence, and numerical errors are treated to improve the accuracy of the results. Principal components decomposition is applied to extract the underlying modes of vibration, showing different characteristics in normal and abnormal phonation. In the last part of this thesis research, the optical flow algorithm for data acquisition as well as the biomechanical model of the vocal fold are used to formulate a nonstationary statistical inverse problem for vocal fold features estimation that accounts for the model uncertainty. An expectation-maximization algorithm for missing data is proposed to find estimates of the system's unknowns. Due to time limitations no computational results are shown and a purely theoretical discussion is presented.

Resumé

Identifikation af abnormiteter på stemmebåndene ved hjælp af forskellige diagnostiske metoder er et særdeles vigtigt trin for at klarlægge årsagen eller årsagerne til problemer med stemmen og derefter give en passende behandling. Til dette formål gavnes kliniske undersøgelser af matematiske modeller til forudsigelse, analyse og estimering. Denne afhandling omhandler biomekaniske modeller af stemmebåndene, særligt af kollisionen og derudover laryngale video-endoskopiske analysemetoder, som er brugbare til afdækning af de tilgrundliggende karakteristika af stemmebåndene.

Den første del af denne forskning omhandler modellering af friktionsløs kontakt under asymmetrisk stemmebåndsvibration. Problemet gribes numerisk an med en tredimensionel “finite element”-model af stemmebåndene med positionsbaserede restriktioner for kontakt. Det er vist, at en ny mekanisme til kontaktdetektion succesfuldt kan detektere kollisioner i asymmetriske oscillationer. Som et alternativ til fjederlignende kontaktkræfter, er optimeringsprocedurer til håndhævelse af restriktioner undersøgt for at forbedre præcisionen af forudsigelserne. Den anden del af denne forskning undersøger et ikke-invasivt indgreb til at kvantitativt analysere den todimensionelle deformation af stemmebåndene optaget med laryngal højhastighedsvideo-endoskopi. En algoritme for tæt optisk strømning er tilpasset billedsekvensens særlige komplekse natur og numeriske fejl behandles for at forbedre resultaternes præcision. Opdeling i hovedkomponenter benyttes til at uddrage de tilgrundliggende bølgetyper i vibrationen, hvilket viser forskellige karakteristika i normal og unormal fonation. Algoritmen for optisk strømning til dataindsamling samt den biomekaniske model af stemmebåndene benyttes i den sidste del af denne afhandling til at formulere et ikke-stationært statistisk inverst problem til estimering af stemmebåndes karakteristiske egenskaber, som tager højde for modellens usikkerhed. En “expectation-maximization”-algoritme til manglende data foreslås til at finde estimater af systemets ubekendte.

Acknowledgements

My primary thanks goes to my supervisor, Jonas Brunskog, who has always supported and encouraged me, giving always constructive comments. I definitely owe him my current way of understanding science. I would like to express my sincere gratitude to Marek Misztal, for his unconditional great help since the very beginning of this thesis, specially for the finite element investigations. I am fully convinced that they possess those human qualities that make the world a better place.

I feel very lucky I ended up by chance in this wonderful and highly altruistic voice production community. Meeting researchers such as Andreas Selamtzis, Scott Thomson, Alexander Lodermeier or Oriol Guasch persuaded me to think so. Also, a big thanks to Andrea, Andreas and Joan who reviewed the full thesis or part of it from a certain distance.

To all of my present and past flat, house, office and lunch-mates, friends here and there, with whom I have had an incredibly fun and invaluable time, thank you. I want to thank my parents, my brothers, and my grandmother for their enormous love, support and company.

Finally, I would like to say that this PhD study made me think a lot about the history of science. I would have liked to share my thoughts with my former supervisor, Finn Jacobsen, and also my grandfather, Ricardo, whose hopes were dashed when his little granddaughter decided not to study medicine. This thesis is dedicated to their memory.

Related publications

During the course of this research, the following papers were produced:

Paper A A numerical strategy for finite element modeling of frictionless asymmetric vocal fold collision. *International Journal for Numerical Methods in Biomedical Engineering*, 2016. doi: [10.1002/cnm.2793](https://doi.org/10.1002/cnm.2793)

Paper B Vocal fold collision modeling. *Proceedings of the 9th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, Italy, 2015.

Paper C Optical flow analysis of high-speed digital imaging of vocal fold vibration. *IEEE Transactions On Biomedical Engineering*. **Submitted**

Paper D Image-based parameter identification of a finite element model of the vocal folds. **Manuscript**

Several other papers were published in the course of the Ph.D., although they are not explicitly cited in this thesis due to overlapped content with the papers A to D.

- Granados, A., Brunskog, J., and Jacobsen, F. (2013). Biomechanical models of damage and healing processes for voice health. *Proceedings of Meetings on Acoustics*, Montréal, Canada.
- Granados, A., Brunskog, J., Misztal, M. K., Visseq, V., and Erleben, K. (2014). Finite element modeling of the vocal folds with deformable interface tracking. *Proceeding of forum acousticum*, Kraków, Poland.
- Granados, A., and Brunskog, J. (2015). Inverse problem in high-speed recordings of the vocal folds. *Book of Abstracts of the 11th Pan-European Voice Conference*, Florence, Italy.
- Granados, A., and Brunskog, J. An optical flow-based state-space model of the vocal folds. *Journal of the Acoustical Society of America*. **Submitted**

Contents

Abstract	i
Resumé	iii
Acknowledgements	v
Related publications	vii
1 Introduction	1
1.1 Motivation	1
1.2 Aims of the thesis	3
1.3 Thesis structure	4
2 Biomechanical models of the vocal folds	7
2.1 Vocal fold anatomy and physiology	8
2.2 Continuum model	9
2.2.1 Fluid	10
2.2.2 Solid	12
2.3 Discretization	17
2.3.1 Spatial discretization	17
2.3.2 Temporal discretization	20
2.4 Numerical solution	24
2.5 Contributions	27
3 High-speed videoendoscopic data acquisition	29
3.1 Optical flow	30
3.1.1 Continuity equation	30
3.1.2 Regularization	30

3.1.3	Robust statistics	34
3.1.4	Aliasing	35
3.2	Principal components analysis	36
3.3	Contributions	38
4	Bayesian statistical inversion	39
4.1	Short introduction to probability	40
4.2	Statistical inversion	41
4.3	Nonstationary inverse problem	44
4.3.1	Filtering problem	45
4.3.2	Smoothing problem	46
4.3.3	State-space identification problem	46
4.4	Contributions	48
5	Overall conclusions	51
5.1	Summary of main results	51
5.2	Directions of further research	53
Paper A	A numerical strategy for finite element modeling of frictionless asymmetric vocal fold collision	59
Paper B	Vocal fold collision modeling	87
Paper C	Optical flow analysis of high-speed digital imaging of vocal fold vibration	93
Paper D	Image-based parameter identification of a finite element model of the vocal folds	103

CHAPTER 1

Introduction

1.1 Motivation

Medicine aims at giving explanations of causes and consequences of health and disease based on empirical evidence, which typically involves detailed examination of the inner body. Application of the experimental method would historically require exploration with the human eye into the anatomy by dissection of corpses (Gray, 1858), or into living structures by vivisection (Bernard, 1865), accompanied by a qualitative analysis. However, these scientific procedures present obvious practical and ethical limitations, which restricted the performance of casual studies in medicine, and in turn motivated the use of alternatives from scientific and technological advances. In the field of study of the human phonatory system, the first evidence for vocal fold physiology studies in living humans were found by means of a laryngeal mirror, termed “auto-laryngoscopy”, for voice pedagogue purposes (Garcia, 1854), which was later introduced into the medical field of laryngology (Czermak, 1861); see Fig. 1.1. However, hypotheses of a vibratory behavior of the vocal folds could not satisfactorily be explained with simple human eye observation. An apparent slow vibration of the vocal folds was first visualized by means of a laryngeal stroboscope (Oertel, 1895), a device which is nowadays widely used in medicine (Bless et al., 1987). Subsequently, evidence of high-frequency vocal fold oscillation were found with high-speed imaging (Farnsworth, 1940; Švec

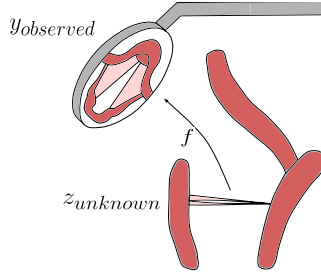


Figure 1.1: Schematic representation of laryngoscopy.

and Schutte, 1996), and laryngeal high-speed videoendoscopy, a procedure able to capture detailed spatial structures and high temporal resolution (Hess and Gross, 1993; Deliyski et al., 2008). Other non-invasive technological measurement techniques, such as the electroglottograph (Fourcin and Abberton, 1971), helped speech pathologist to work on hypotheses of damage related to vocal fold contact. Furthermore, the need of making predictions of a process to test hypothesis promotes the collaboration between medicine, science and technology, but also mathematics, for further clinical progress. Scientific hypotheses of the vocal fold contact mechanism formulated in terms of mathematical models (Gunter, 2003; Tao et al., 2006) can be tested with *ex-vivo* experiments that make use of pressure sensors (Jiang and Titze, 1994), the results having direct implications in voice health investigations (Titze et al., 2003; Gunter, 2003). In addition, the application of mathematics plays an important role in quantitative analysis of the observations (Wittenberg et al., 1995), and in feature inference related to medical hypothesis of damage that may not be testable in *in vivo* observations. This is a common situation in studies on vocal fold tissue-related injury (Titze et al., 2003), where mathematical optimization is used to infer mechanical parameters from quantitative analysis of laryngeal high-speed videoendoscopy (Döllinger et al., 2002; Cataldo et al., 2013). This particular problem would correspond to find an unmeasurable variables $z_{unknown}$ from measurements $y_{observed}$ via the inverse mapping f^{-1} . That is,

$$z_{unknown} \xleftarrow{f^{-1}} y_{observed} \quad , \quad (1.1)$$

which is illustrated in Fig. 1.1. This formulation actually models the natural process of inference that humans make on a regular basis. For instance, imagine you see someone who laughs ($y_{observed}$) and you want to infer her or his emotional state ($z_{unknown}$). At first approach, you could state that this person is happy. However, it could also be considered that this person laughs because she or he is under an altered mental state. In that situation, the observation is not sufficient to determine the emotional state of this person, or in mathematics,

the function f is non-injective and the inverse f^{-1} does not exist. The system defined by f is then underdetermined, and possible solutions are highly uncertain. However, the uncertainty can be drastically reduced if you observe further that the person who laughs holds a wad of cash, that is, if the number of observations is increased. In addition, if someone is tickling the one who laughs, then the problem becomes more complex as this person adds confusion or noise to the available observations. In mathematics the function f is then called surjective, and in order to find the solution to the problem it is important to identify and exclude the unwanted data. Finally, it could also happen that you know personally the person who laughs, that is, you may have prior knowledge on her or his previous emotional states. If the person who laughs is in general a happy and positive person, you may conclude with a high degree of certainty that the one who laughs is simply happy. However, this prior knowledge that you may have is in turn uncertain, as it is not the same knowing someone from work or from childhood.

In the case of vocal fold features inference, that is, estimation of $z_{unknown}$ from observed laryngeal high-speed videoendoscopy $y_{observed}$, the problem has a similar degree of complexity as the one discussed above: videoendoscopic data may not be sufficient to capture the complexity of the underlying vocal fold dynamics (f injective), qualitative analysis of the data typically contains noise (f surjective), and prior knowledge on the vocal fold dynamics is available via mathematical models, which as such are reliable to some degree of probability. While deterministic inference methods draw conclusions that are logical consequences of the premises, and ignore the actual uncertainty of observation-based knowledge, the mathematical Bayesian formulation (Kaipio and Somersalo, 2005; Cataldo et al., 2013) of the inference problem in Eq. (1.1) captures the uncertainties via Bayes' theorem

$$\pi(z_{unknown}|y_{observed}) = \frac{\pi(z_{unknown})\pi(y_{observed}|z_{unknown})}{\pi(y_{observed})}, \quad (1.2)$$

where $\pi(\cdot|\cdot)$ refers to conditional probability density function. The conclusions that can be drawn from these mathematical models are probable, similarly to the conclusion made from inductive reasoning typical in natural sciences or medicine. Thereby, the mathematical theory for inference provides a mathematical framework suitable for further medical argumentation on the processes of damage.

1.2 Aims of the thesis

Rather than argumentative, this thesis is an analytical study of the inference problem of estimating vocal fold features from laryngeal high-speed endoscopy

with the main purpose of providing complementary mathematical tools for medical scientific investigations on the processes of damage. The problem is divided into three sub-problems concerned with prediction biomechanical models ($z_{unknown}$), laryngeal data acquisition ($y_{observer}$) and inversion formulation (f^{-1}), in Eq. (1.1). These problems are described in three separate chapters (2,3,4), respectively.

The three sub-problems in which the inference problem is based have the following goals (with corresponding chapters):

- Improve accuracy and precision of existing biomechanical models of the vocal folds for inversion purposes, in relation to collision models, optimization procedures and temporal discretization (Chapter 2).
- Provide mathematical tools for quantitative analysis of laryngeal high-speed videoendoscopic observations for inversion purposes, in relation to dense data acquisition and hidden pattern identification (Chapter 3).
- Provide an adequate inverse formulation for vocal fold features estimation, in relation to nonstationary statistical inverse methods (Chapter 4).

1.3 Thesis structure

This PhD dissertation is based on a collection of four scientific papers, included at the end of this thesis. Each chapter contains the mathematical background which sustains the scientific contributions of this thesis. At the end of each chapter, a summary of the scientific contributions in relation to the state of the art is presented. The content of this thesis can be summarized as follows:

- Chapter 2 contains general concepts of continuum solid and fluid mechanics for vocal fold modeling, and spatial and temporal discretization considerations of the finite-element method. The collision problem is introduced, and general optimization techniques to find solutions of the equations of motions are presented. The reader is then referred to Paper A and Paper B.
- Chapter 3 presents some general concepts on image processing techniques. An algorithm to extract two-dimensional dense tissue velocity estimates is presented, and principal components decomposition is discussed as a video analysis tool. The reader is then referred to Paper C.
- Chapter 4 presents mathematical considerations on statistical inverse problems in connection to prediction models (Chapter 2) and measurement data (Chapter 3). The difference between stationary and nonsta-

tionary inversion is described. The reader is then referred to Paper D, where use is made of the results in Paper A-Paper C.

- Chapter 5 summarizes the major conclusions of the three sub-problems and presents some directions for future research.

CHAPTER 2

Biomechanical models of the vocal folds

In physical modelling, the model purpose dictates its design and implementation. In the studies concerned with the vocal fold system, different biomechanical models were created to be used as a simple voice source for speech synthesis (Ishizaka and Matsudaira, 1972), to analyze normal and pathological vocal fold vibration involving the layered vocal fold structure (Story and Titze, 1995), or to create natural sound generation in voice production involving advanced aeroacoustic mechanisms (Šidlof et al., 2015). The purpose of the vocal fold model investigated in this thesis is to provide tools for clinical studies. More specifically, the purpose is twofold: to provide a suitable framework to formulate an inverse problem for vocal fold parameter estimation and to improve the mechanical and dynamical predictions during vocal fold collision. The two aspects, however, are related. As it will be shown in this thesis, the model requires a fine compromise between the airflow and the tissue models, and attention must be paid to accuracy and precision of its predictions.

This chapter covers a review of some structural and mathematical aspects needed to formulate and numerically implement a biomechanical model of the vocal fold with regard to continuum physics, discretization algorithms and solution methods. Standard textbooks in which most of this chapter is based are (Titze, 2006; Lautrup, 2004; Hughes, 2000; Wriggers, 2002). A summary of the scientific contributions of the thesis are presented at the end of the chapter in

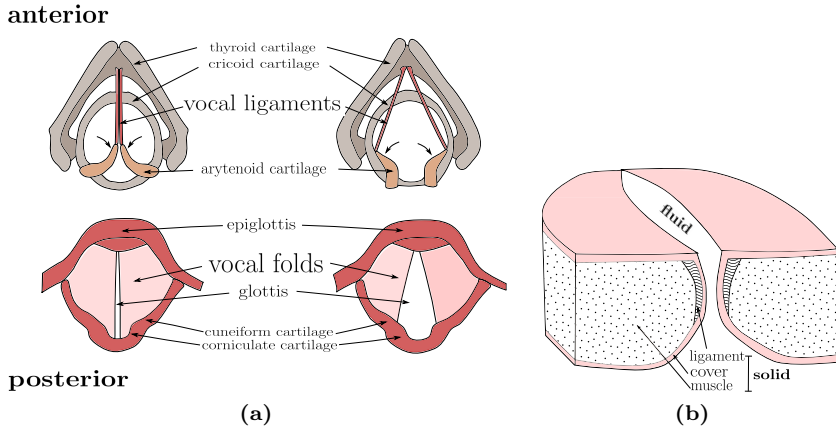


Figure 2.1: (a) Transverse view of the larynx: the interior cartilaginous structure (top), and the exterior mucous membrane (bottom). Two different vocal fold posturing are shown: adduction typically related to phonation or protection of the airway (left), and abduction typically related to breathing (right). (b) Frontal view of the vocal folds with the interior tissue layers.

relation to the state of the art.

2.1 Vocal fold anatomy and physiology

The epiglottis is the cartilage that controls the opening and closure of the entrance to the larynx. The larynx, situated in the neck, is the organ that contains the vocal folds, which are the main source of sound in phonation and are protected by the thyroid cartilage. The anterior prominence of the thyroid cartilage is visible, specially in men, and it is called Adam's apple. The vocal folds are bounded interiorly by the airflow domain, or glottis, laterally and anteriorly by the thyroid cartilage, and inferiorly by the ring-shaped cricoid cartilage; see Fig. 2.1a for an illustration. The vocal folds are attached posteriorly to the base of the pyramidal-shaped arythenoid cartilages, at the vocal processes, and anteriorly to the thyroid cartilage via the vocal fold ligaments. On top of the arythenoid cartilages, the corniculate cartilages and the wedge-shaped cuneiform cartilages are found.

Phonation is controlled by the activation of a number of intrinsic laryngeal muscles, connected to the above-mentioned cartilages. The mechanical interaction of these cartilages results into large and slow deformations that define different posturing of the vocal folds (adduction, abduction or elongation) characteristic of, e.g, the speaking voice, singing voice, loud voice, or breathy voice. Given a

vocal fold position for voiced sound production, the subglottal pressure generated by the lungs create an upward aerodynamic fluid flow which interacts with the soft tissue of the vocal folds, setting the latter into passive self-oscillation (Lucero et al., 2009). Most of the deformation occurs in the superior mucous membrane or cover, formed by the superficial layer of the lamina propria and the epithelium, that has very flexible properties. A wave-like motion of the vocal folds takes place from the bottom to the top of the vocal folds. The thyroarytenoid muscle or body, which forms the innermost layer of the vocal folds, remains mostly static as well as the intermediate and deep layers of the lamina propria that comprise the vocal ligament. A simplified schema of the vocal fold's inner structure is shown in Fig. 2.1b.

This thesis is concerned with the process of passive self-oscillation of the vocal folds. Therefore, vocal fold posturing dynamics is disregarded.

2.2 Continuum model

In fact, the vocal fold is composed of multiple tissue layers with different properties that arise from the interaction of microscopic variables. Multimass models of the vocal fold (Ishizaka and Flanagan, 1972; Story and Titze, 1995; Yang et al., 2010) represent the tissue as a collection of a reduced number of point particles with a fixed mass, which interact according to Newton's laws. The dynamics of the vocal fold is then described as a number of coupled ordinary differential equations. In applications where higher precision is desired, such as in clinical studies where detailed pathological tissue variations are of importance, the number of point particles must be very large in order to capture the apparent smoothness of the observed vocal fold. However, this results into a large number of systems with parameters difficult to control. Alternatively, a smooth continuum description of the vocal fold can be obtained in the framework of continuum physics. A continuum model of the vocal fold consists of material particles with physical properties characterized by smooth functions of space and time, such as mass density, to which Newton's laws apply.

Because laryngeal high-speed videoendoscopy is able to capture the complex superior geometries and deformations, a continuum model of the tissue (Titze and Talkin, 1979) is here developed. In addition, continuum models open the possibility to further investigate intricate phenomena that occur during self-oscillation, such as collision (Paper A).

The vocal fold oscillation cause a rapid compression of the air particles placed within the glottal area. Then, glottal pulses generate an acoustic wave that travels upwards through the vocal tract, which acts as a filter, before it is radiated

from the mouth, reaching atmospheric pressure conditions. However, the models used in the thesis do not aim at predicting the acoustic consequences, therefore vocal tract coupling is disregarded. This would require a drastic increase of the model complexity by including non-linear acoustic-structure interaction (Titze, 2008; Šidlof et al., 2015), specially when laryngeal observations are taken with laryngoscopic devices that obstruct the natural airflow.

In the absence of volume forces, the equations of motion presented in this chapter are derived from the equilibrium equations

$$\nabla \cdot \boldsymbol{\sigma} = \rho \frac{D\mathbf{v}}{Dt} \quad (2.1)$$

where \mathbf{v} represents the particle velocity of the medium, ρ the density of the medium and

$$\frac{D\mathbf{v}}{Dt} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \quad (2.2)$$

the material derivatives, which contain the temporal and advection terms, referring to the temporal rate of change at a particular position and the spatial rate of change at a particular time instant. The tensor $\boldsymbol{\sigma}$ describes the force per unit area or stress state in a given particle

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix}, \quad (2.3)$$

where the diagonal elements correspond to the normal (pressure) stresses, and the non-diagonal refer to the shear (viscous) stresses. The force acting on an arbitrary surface element satisfies $d\mathbf{f} = \boldsymbol{\sigma} \cdot d\mathbf{s}$. The total is therefore

$$f_i = \oint \sum_j \sigma_{ij} ds_j, \quad (2.4)$$

which, after applying Gauss' theorem, results into the left-hand side term in Eq. (2.1).

2.2.1 Fluid

The air pressure acting on the vocal fold surface has a strong influence on the vibratory pattern of the vocal folds (Van Den Berg et al., 1957). Glottal aerodynamics can be expressed by the full Navier-Stokes equations containing turbulent, viscous, advective terms and assuming compression. In this thesis, a simplified model of the flow is prioritized for two reasons. First, as it will

be shown later in Chapter 4, inversion procedures require noticeable manipulations of system matrices. Because the use of continuum models increases the complexity of the system, a full fluid-structure interaction model of the vocal fold would result into an intractable inverse problem. Second, the observation data from laryngeal videoendoscopy can only capture the vocal fold motion, and disregards pressure measures. Then, an inference problem including an advanced fluid model would carry large uncertainties. Therefore, a number of simplifications are presented in this section, supported by multiple theoretical studies and *in-vitro* experiment (e.g., see Scherer et al., 2001; Cisonni et al., 2008; Van Hirtum et al., 2009; Mittal et al., 2013).

Reynolds numbers, that is, the ratio of the advective to the viscous terms, of the fluid flow within the glottis have values $Re = \mathcal{O}(10^3) \gg 1$. These moderate values support the assumption of an inviscid flow. Low Mach numbers, that is, the ratio of the flow particle velocities to the speed of sound in air c , are found to be $Ma^2 = \mathcal{O}(10^{-2}) \ll 1$. Thus, the flow can be assumed to be effectively incompressible because

$$\left| \oint_S \mathbf{v} \cdot d\mathbf{S} \right| = |\nabla \cdot \mathbf{v}| \leq \frac{\mathbf{v}^2}{c^2} |\nabla \mathbf{v}| = Ma^2 |\nabla \mathbf{v}|, \quad (2.5)$$

which indicates that the instantaneous rate of change of a volume matter bounded by S is very small. Finally, Strouhal numbers, that is, the ratio of the effective diameter of the object to the velocity, are found to be $Sr = \mathcal{O}(10^{-2}) \ll 1$, which supports the assumption of a steady flow. The equilibrium equations in Eq. (2.1) are therefore the Euler equations for a steady and incompressible flow

$$(\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{1}{\rho} \nabla p, \quad (2.6a)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (2.6b)$$

By integrating the Euler equations along streamlines, the equilibrium equations are equivalent to Bernoulli's theorem for incompressible fluid, that states that the functional

$$H = \frac{1}{2} \mathbf{v}^2 + \frac{p}{\rho} \quad (2.7)$$

is constant along streamlines (Lautrup, 2004); see Fig. 2.2. This means that an increase in the particle velocity in contracting canals must be compensated by a decrease in the pressure. It is assumed further that the flow is one-dimensional along the inferior-superior dimension. This one-dimensional Bernoulli flow can be used to find simple estimates of the pressure within the glottal constriction in a continuum model of the vocal folds (Paper A).

These simplifications result into some limitations. For expanding junctions, as at the glottal exit, the flow detaches the boundary layer (separation point).

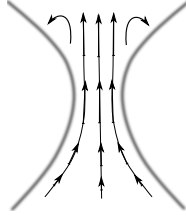


Figure 2.2: Schema of the fluid flow through a constriction.

Then, downstream the separation point the flow can reverse in turbulences and in the central part a jet is formed (Pelorson et al., 1994). Because the simplified one-dimensional Bernoulli flow neglects viscous forces necessary to form a laminar boundary layer and cannot successfully predict the dynamics at the glottal exit, experimental results are then used to predict the exit pressure recovery (Ishizaka and Matsudaira, 1972). Similarly, the flow detachment point cannot be predicted by an inviscid one-dimensional flow model. Some approximations will be presented in Sec. 2.3.1.

2.2.2 Solid

Under three-dimensional vocal fold tissue deformation a solid particle is displaced from its original position (material coordinates) $\mathbf{X} = (X_1, X_2, X_3) \in V_{\text{solid}}$ to its current position (spatial coordinates) $\boldsymbol{\varphi}(\mathbf{X}) = \mathbf{x} = (x_1, x_2, x_3) \in v_{\text{solid}}$ via the deformation mapping $\boldsymbol{\varphi} : V_{\text{solid}} \rightarrow \mathbb{R}^3$. This can be described by the displacement field

$$\mathbf{u} \equiv \mathbf{u}(\mathbf{X}, t) = (u_1(\mathbf{X}, t), u_2(\mathbf{X}, t), u_3(\mathbf{X}, t)) = \mathbf{x} - \mathbf{X} \quad (2.8)$$

This is called the Lagrange representation, in contrast to the Euler representation $\mathbf{u}(\mathbf{x}, t)$. As it will be shown later in Sec. 2.3.1, the use of a Lagrange representation drastically simplifies the calculations. The current position can hence be written as $\mathbf{x} = \mathbf{X} + \mathbf{u}(\mathbf{X}, t)$. Because the material coordinate \mathbf{X} is time-independent, the particle velocity is $\mathbf{v} \equiv \mathbf{v}(\mathbf{X}, t) = \frac{\partial \mathbf{u}}{\partial t}$ and the acceleration $\frac{\partial^2 \mathbf{u}}{\partial t^2}$. Then, from the equilibrium equations in Eq. (2.1), the governing equations of motion of the vocal fold tissue are described by

$$\nabla \cdot \boldsymbol{\sigma} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad \text{in } v_{\text{solid}}, \quad (2.9)$$

where ρ is the spatial mass density. These equations are known as Navier's equations of motion. The stress tensor $\boldsymbol{\sigma}$ of the tissue is derived in what follows.

The deformed differential $d\mathbf{x}$ can be written as

$$dx_i = \sum_j \frac{\partial x_i}{\partial X_j} dX_j \quad (2.10)$$

for the i -th coordinate, where $\frac{\partial x_i}{\partial X_j}$ is the Jacobian of the transformation $\boldsymbol{\varphi}$

$$\mathbf{F} = \frac{\partial \boldsymbol{\varphi}(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} = \frac{\partial \mathbf{u}}{\partial \mathbf{X}} + \mathbf{I}$$

known as the deformation gradient. The length of the deformed differential is

$$d\mathbf{x}^2 = \sum_{jk} g_{jk} dX_j dX_k \quad (2.11)$$

where the coefficients g_{jk} are

$$g_{jk} = g_{jk}(\mathbf{X}) = \sum_i \frac{\partial x_i}{\partial X_j} \frac{\partial x_i}{\partial X_k} \quad (2.12)$$

The expression above can be written as

$$g_{ij} = \delta_{ij} + 2E_{ij}, \quad (2.13)$$

where

$$E_{ij} = \frac{1}{2} \left(\nabla_i^0 u_j + \nabla_j^0 u_i + \sum_k \nabla_i^0 u_k \nabla_j^0 u_k \right) \quad (2.14)$$

is the Green's strain tensor

$$\mathbf{E} = \frac{1}{2} (\mathbf{F}^T \mathbf{F} - \mathbf{I}), \quad (2.15)$$

and the super-index 0 indicates derivatives in terms of material coordinates. If small displacement gradients are assumed (smooth displacement field), then $|\nabla_j^0 u_i| \approx |\nabla_j u_i| \ll 1$. It is assumed further that the displacement field is small compared to the size of the vocal folds. It is important to consider these assumptions because they drastically simplify the implementation of the discrete model, discussed later in Sec. 2.3. In addition, the material and spatial derivatives can then be interchangeable. Under these assumptions, the strain tensor is called Cauchy's strain tensor and it is written as

$$\epsilon_{ij} = \frac{1}{2} (\nabla_i u_j + \nabla_j u_i) \quad (2.16)$$

The tensor is symmetric and a common notation is $\boldsymbol{\epsilon}$.

Whenever external forces, such as the intraglottal aerodynamic pressure, are applied to the body, internal forces are set up described by the symmetric Cauchy stress tensor in the deformed configuration v_{solid} . In linear elasticity theory, Hooke's law implies that the stress is a linear function of the strain. At first approximation, the vocal fold tissue is assumed to obey Hooke's law under a small deformation regime. The Young's modulus E expresses the linear relation between stress and strain in the same direction, and the Poisson's ratio ν expresses contraction in a direction transverse to the direction of extension. These stress-strain relations must be of course independent of the coordinate system. In an isotropic material, i.e., a material which has the same properties in all directions, a relation invariant under change of coordinates is

$$\sigma_{ij} = 2\mu\epsilon_{ij} + \lambda \sum_k \epsilon_{kk} \delta_{ij} \quad (2.17)$$

where λ and μ are the Lamé coefficients. The latter is called shear modulus, also denoted by G in the literature. The Young's modulus and Poisson's ratio can be written as a function of the Lamé coefficients. However, the vocal fold tissue is assumed to have transverse isotropy, that is, material properties in the anterior-posterior direction (y -direction along the fiber) differ from material properties in the frontal plane (xz -plane). Being (E, ν) and $\mu = \frac{E}{1+\nu}$ the material properties in the frontal plane and (E', ν', μ') in the anterior-posterior direction, then the total strain in the x direction is the summation of all deformations in that direction. All the relations form the constitutive equations

$$\begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \sigma_{xy} \\ \sigma_{yz} \\ \sigma_{zx} \end{pmatrix} = \mathcal{S} \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ 2\epsilon_{xy} \\ 2\epsilon_{yz} \\ 2\epsilon_{zx} \end{pmatrix} \quad (2.18)$$

where (Berry and Titze, 1996)

$$\mathcal{S} = \begin{pmatrix} \frac{(E' - \nu'^2 E)E}{(1+\nu)k} & \frac{EE'\nu'}{k} & \frac{(\nu'^2 E + \nu E')E}{(1+\nu)k} & 0 & 0 & 0 \\ \frac{EE'\nu'}{k} & \frac{E'^2(1-\nu)}{k} & \frac{EE'\nu'}{k} & 0 & 0 & 0 \\ \frac{(\nu'^2 E + \nu E')E}{(1+\nu)k} & \frac{EE'\nu'}{k} & \frac{(E' - \nu'^2 E)E}{(1+\nu)k} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu' & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu' & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix} \quad (2.19)$$

is the stiffness matrix and $k = E'(1 - \nu) - 2\nu'^2 E$. From thermodynamic argumentation, it can be seen that (Lempriere, 1968)

$$\nu < 1 - 2\nu'^2 E/E' \quad (2.20)$$

When different tissue layers travel at different speeds, viscous shear forces arise which are defined from gradients of velocities as

$$\sigma_{ij} = \eta \frac{\partial v_i}{\partial x_j} \quad (2.21)$$

controlled by the parameter η which is called dynamic viscosity or viscosity. These viscous terms arise from the rate of change of the strain $\dot{\epsilon}(t)$

$$\frac{\partial}{\partial t} \epsilon_{ij}(t) = \frac{\partial}{\partial t} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) = \frac{\partial}{\partial x_j} \frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_i} \frac{\partial u_j}{\partial t} = \frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \quad (2.22)$$

Then, the tissue constitutive equation in Eq. (2.18) is can be modeled as

$$\boldsymbol{\sigma}(t) = \mathcal{S}\boldsymbol{\epsilon}(t) + \eta\dot{\boldsymbol{\epsilon}}(t) \quad (2.23)$$

This corresponds to a Kelvin-Voigt model, similar to a spring and a damper in parallel. In terms of creep and relaxation responses, this model reflects the following soft tissue behavior. The creep response is the strain function in time for a given constant stress $\sigma(0)$. In the one-dimensional case, the strain response solution is shown in Fig. 2.3(left). The nonlinear bounded creep by the Kelvin-Voigt element provide realistic description of the commonly observed behavior of soft tissues. Similarly, the relaxation response is the stress function in time for a given constant strain $\epsilon(0)$, and the solution is shown in Fig. 2.3(right). The model shows a lack of relaxation which makes it unrealistic, and a limiting factor in the modeling of the tissue. Alternative models, such as the standard linear model consisting on a damper and a spring in parallel connected with a second spring in series, may provide a potential improvement in vocal fold models, as well as hyperelastic tissue models. However, existing studies in this direction consider large deformations in vocal fold posturing and the consequent stretch for fundamental frequency control (Zhang et al., 2006) as an improvement to classical string models (Titze, 1989). As mentioned in Sec. 2.1, this thesis concerns with passive self-oscillation under a small displacements regime. Hence, the gain obtained with the use of a hyperelastic model during vocal fold self-oscillation is not evident (Bakhshaei, 2013). However, it has been shown that vocal fold function is related to the linear viscoelasticity of the vocal fold tissue (Titze, 1988), specially in the vibratory capacity of the vocal fold cover (Chan et al., 2001).

Along with the Navier's equations in Eq. (2.9), a number of physical constraints must be taken into account. Firstly, the vocal fold tissue is attached to the thyroid cartilage and to the base of the arythenoid cartilage (vocal fold process). Assuming that the tissue has zero displacement in the region of attachment denoted as $\Gamma_D \subset \partial v_{solid}$, then a Dirichlet boundary condition can be enforced as

$$\mathbf{u} = 0 \quad \text{on } \Gamma_D \quad (2.24)$$

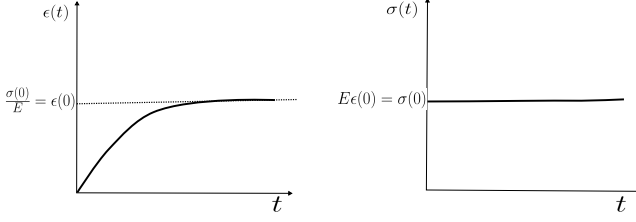


Figure 2.3: Creep (left) and relaxation (right) response of a Kelvin-Voigt viscoelastic tissue model.

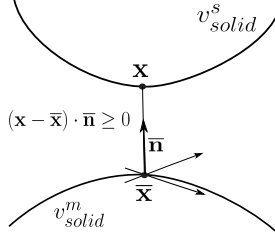


Figure 2.4: Non-penetration contact constraint.

In addition, at the intraglottal vocal fold boundary $\Gamma_N \subset \partial v_{solid}$, the aerodynamic pressure acts as an external force that sets the system into oscillation, which can be modeled by the Neumann boundary

$$\boldsymbol{\sigma} \cdot \mathbf{n} = \mathbf{p} \quad \text{on } \Gamma_N, \quad (2.25)$$

where \mathbf{n} is the outward normal on the boundary Γ_N . Furthermore, when vocal fold collision occurs, additional boundary conditions on the contact surface $\Gamma_C \subset \partial v_{solid}$ arise to avoid physically observed body interpenetration (Paper A). A non-negative normal gap is enforced as

$$(\mathbf{x} - \bar{\mathbf{x}}) \cdot \bar{\mathbf{n}} = (\mathbf{u} - \bar{\mathbf{u}}) \cdot \bar{\mathbf{n}} + (\mathbf{X} - \bar{\mathbf{X}}) \cdot \bar{\mathbf{n}} \geq 0 \quad \text{on } \Gamma_C. \quad (2.26)$$

This constraint is illustrated in Fig. 2.4, where the particle \mathbf{x} placed in a body v^s_{solid} is prevented to penetrate into the body v^m_{solid} , and $\bar{\mathbf{x}}$ is the projection of \mathbf{x} onto this surface, that is,

$$\bar{\mathbf{x}} = \arg \min_{\mathbf{y} \in \partial v^m_{solid}} \{\|\mathbf{x} - \mathbf{y}\|_2\} \quad (2.27)$$

It is necessary to assume that v^m_{solid} is a convex body and ∂v^m_{solid} is smooth (Wriggers, 2002). Otherwise, uniqueness and existence problems of the projection point can occur, as will be described in the following section.

2.3 Discretization

In order to computationally solve and simulate a complex continuum model with the aim to make useful predictions, numerical methods must be applied (Hughes, 2000). To this end, the continuum set of points assumed in the space and time domains must be discretized by discretization methods.

2.3.1 Spatial discretization

Fuid

By spatial discretization of the air canal into cross sectional areas $a(z_i)$ in the inferior-superior direction (z), the steady one-dimensional Bernoulli flow in Eq. (2.28) yields

$$\begin{aligned} \frac{1}{2}v(z_i)^2 + \frac{p(z_i)}{\rho} &= \frac{1}{2}v(z_{i+1})^2 + \frac{p(z_{i+1})}{\rho} \\ \Rightarrow p(z_{i+1}) &= p(z_i) + \frac{\rho}{2}U^2 \left(\frac{1}{a(z_i)^2} - \frac{1}{a(z_{i+1})^2} \right), \end{aligned} \quad (2.28)$$

where $v(z_i)$ are average flow speeds defined as

$$v(z_i) = \frac{U}{a(z_i)} = \frac{1}{a(z_i)} \int_{a(z_i)} \mathbf{v}(z_i) \cdot d\mathbf{s}$$

and $U \equiv U(t)$ is the volumetric flow rate. Back to Euler's equations in Eq. (2.6) and given its discretization, it can be seen that the pressure drop from the point of flow detachment z_d and the epiglottal region z_e can be approximated as (Titze, 2006)

$$p(z_d) - p(z_e) = -\rho v(v(z_d) - v(z_e)), \quad (2.29)$$

where v is an intrajunction velocity which is assumed to have a parametric form as

$$v = \frac{1}{2}(\alpha v(z_d) + (2 - \alpha)v(z_e)) \quad (2.30)$$

where α is a parameter to be estimated experimentally. Setting $\alpha = 1$, the flow at the expanding junction behaves as a Bernoulli flow, and no energy loss is estimated. Intermediate values of α may depend on the vocal fold characteristics. Furthermore, at the glottal entrance the area contraction is noticeable, and the subglottal kinetic energy can be neglected. In addition, it is typically assumed that the subglottal pressure is equal to a constant lung pressure, in the absence of acoustic waves. Finally, the flow separation point z_d must be estimated. The simplest model assumes a jet formation upstream of the minimum glottal area (Paper A), although more complex models can be used.

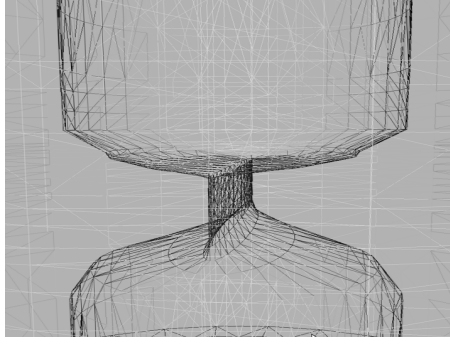


Figure 2.5: Spatial discretization of the larynx.

Solid

The finite element method is applied to find the vocal fold displacements in the spatially discretized vocal fold structure. The starting point of the method is the weak form of Galerkin of Eq. (2.9). Multiplication at both sides by admissible test functions \mathbf{w} in v_{solid} vanishing in the Dirichlet domain Γ_D , integration over the current solid domain v_{solid} and application of Gauss' divergence theorem yields the weak form (Hughes, 2000)

$$\int_{v_{solid}} \boldsymbol{\sigma} : \nabla \mathbf{w} \, dv + \int_{v_{solid}} \rho \frac{\partial \mathbf{u}^2}{\partial t^2} \cdot \mathbf{w} \, dv = \int_{\Gamma_N} \mathbf{w}^T \cdot \mathbf{p} \, d\Gamma, \quad (2.31)$$

where use is made of the equality $\nabla \cdot (\boldsymbol{\sigma} \cdot \mathbf{w}) = \boldsymbol{\sigma} : \nabla \mathbf{w} + \nabla \cdot \boldsymbol{\sigma} \cdot \mathbf{w}$. The problem is now to find a solution \mathbf{u} satisfying the Dirichlet boundary conditions such that the weak form in Eq. (2.31) is satisfied for all test functions \mathbf{w} . For the sake of computational simplicity and given the small and smooth displacement assumptions discussed earlier in the chapter, it is possible to define test functions in the material coordinates and compute integrals over the material volume V_{solid} . Using the stress-strain constitutive equations, the weak form is

$$\begin{aligned} \int_{V_{solid}} \rho \mathbf{w}^T \cdot \frac{\partial \mathbf{u}^2}{\partial t^2} \, dV + \int_{V_{solid}} \boldsymbol{\epsilon}(\mathbf{w})^T \cdot \mathcal{S} \cdot \boldsymbol{\epsilon}(\mathbf{u}) \, dV + \int_{V_{solid}} \boldsymbol{\epsilon}(\mathbf{w})^T \cdot \boldsymbol{\eta} \cdot \boldsymbol{\epsilon} \left(\frac{\partial \mathbf{u}}{\partial t} \right) \, dV \\ = \int_{\Gamma_N} \mathbf{w}^T \cdot \mathbf{p} \, d\Gamma \end{aligned} \quad (2.32)$$

Notice that the external force component on the left-hand side of the equation is here computed in the spatial coordinates, for computational simplicity. By

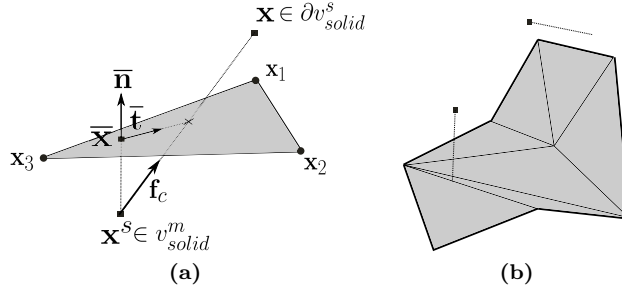


Figure 2.6: (a) Node-to-surface contact discretization. (b) Node-to-surface contact detection where some colliding nodes are placed in blind spots.

decomposition of the admissible test functions \mathbf{w} in a finite number of linear shape functions defined in a tetrahedral mesh illustrated in Fig. 2.5, the finite element system matrices can be easily derived by means of isoparametric transformations, leading to a system of the form (Paper A)

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{F}, \quad (2.33)$$

where \mathbf{M} , \mathbf{C} , and \mathbf{K} are the mass, damping, and stiffness matrices and \mathbf{F} is the vector of applied aerodynamic forces. Because the material coordinates are time-independent, the system matrices are time-independent and are computed only once.

When collision occurs, two finite elements penetrate to each other and a contact discretization procedure must be performed to enforce necessary constraints that avoid volume interpenetration. The simplest form of discretization is node-to-node, consisting of selecting matching nodes of a colliding body mesh. This formulation is valid for a symmetric and conforming mesh, and when small displacements are assumed (Wriggers, 2002). However, a three-dimensional finite element mesh of the vocal folds may show asymmetries and intricate geometries that make the predefinition of nodes complicated, specially for abnormal phonation modeling. In addition, this formulation does not allow for sliding between contact surfaces, as the tangent component $\bar{\mathbf{t}}$ of the reaction contact force \mathbf{f}_c cannot be defined. As a solution, a node-to-surface contact element can be defined, which allows for sliding contact and asymmetric geometries in non-conforming meshes (Paper A). In Fig. 2.6a an illustration of the contact element is shown, where a node placed prior to collision at $\bar{\mathbf{x}}$ is displaced to the collision position \mathbf{x}^s . The triangular interface geometry can be easily parametrized by an isoparametric transformation with linear functions $N_i(\xi, \zeta)$ defined in a local

coordinate system. Then, the projection point in Eq. (2.27) is

$$\bar{\mathbf{x}} = \sum_{i=1}^3 N_i(\bar{\xi}, \bar{\nu}) \mathbf{x}_i \quad (2.34)$$

with

$$\begin{pmatrix} \bar{\mathbf{a}}_1 \cdot \bar{\mathbf{a}}_1 & \bar{\mathbf{a}}_1 \cdot \bar{\mathbf{a}}_2 \\ \bar{\mathbf{a}}_2 \cdot \bar{\mathbf{a}}_1 & \bar{\mathbf{a}}_2 \cdot \bar{\mathbf{a}}_2 \end{pmatrix} \cdot \begin{pmatrix} \bar{\xi} \\ \bar{\nu} \end{pmatrix} = \begin{pmatrix} (\mathbf{x}^s - \mathbf{x}_1) \bar{\mathbf{a}}_1 \\ (\mathbf{x}^s - \mathbf{x}_1) \bar{\mathbf{a}}_2 \end{pmatrix} \quad (2.35)$$

and

$$\bar{\mathbf{a}}_1 = (\mathbf{x}_2 - \mathbf{x}_1) \quad \bar{\mathbf{a}}_2 = (\mathbf{x}_3 - \mathbf{x}_1)$$

The penetration distance $(\mathbf{x}^s - \bar{\mathbf{x}}) \cdot \bar{\mathbf{n}}$ in Eq. (2.26) can then be computed.

Furthermore, the node-to-surface contact elements have to be identified in order to formulate the constraint. Contact detection methods based on the penetration distance require to identify at a time step immediately prior to collision the node and surface pair that are likely to collide. A classical criterion says that if the distance in Eq. (2.27) between the node and the closest surface, found in a combinatorial search, is below a given threshold, then contact may occur. However, for intricate geometries, nodes can be placed in blind spots, where the projection point does not exist. In Fig. 2.6b two examples are shown: a node that cannot project onto an edge, and a node located too far from the solid domain. The first situation could be solved by smoothing the contact surface with, e.g., splines or Bézier interpolation, at the expense of increasing algebraic and computational complexity (Wriggers, 2002). Alternatively, the two problems could be solved by defining a different contact detection criterion (Paper A).

2.3.2 Temporal discretization

The temporal discretization of the spatially discretized equilibrium equations is approached by a direct time-integration method based on the finite difference formulation

$$\mathbf{u}_{n+1} = \sum_{j=1}^m \alpha_j \mathbf{u}_{n+1-j} - h \sum_{j=0}^m \beta_j \dot{\mathbf{u}}_{n+1-j}, \quad (2.36)$$

with time step $h = t_{n+1} - t_n$. Algorithms belonging to the Newmark's family are typically used to solve the equilibrium equations such as the vocal fold equations of motion. These methods are single-step ($\alpha_j = \beta_j = 0$ for $j > 1$), that is, the solution depends on the solution at the immediately preceding time; this is an important characteristics for the simplification of a statistical inverse problem

(Paper D). The Newmark method can be derived from Taylor series expansions, so that the velocities and positions can be approximated as (Gérardin, 1994)

$$\dot{\mathbf{u}}_{n+1} = \dot{\mathbf{u}}_n + \int_{t_n}^{t_{n+1}} \ddot{\mathbf{u}}(\tau) d\tau \quad (2.37a)$$

$$\mathbf{u}_{n+1} = \mathbf{u}_n + h\dot{\mathbf{u}}_n + \int_{t_n}^{t_{n+1}} (t_{n+1} - \tau) \ddot{\mathbf{u}}(\tau) d\tau, \quad (2.37b)$$

where the integrals correspond to the truncation error, and the subscript n indicates that the function is approximated at time t_n . From the approximations in the time interval $[t_n, t_{n+1}]$

$$\begin{aligned} \ddot{\mathbf{u}}_n &= \ddot{\mathbf{u}}(\tau) + \mathbf{u}^{(3)}(\tau)(t_n - \tau) + \mathbf{u}^{(4)}(\tau)\frac{(t_n - \tau)^2}{2} + \dots \\ \ddot{\mathbf{u}}_{n+1} &= \ddot{\mathbf{u}}(\tau) + \mathbf{u}^{(3)}(\tau)(t_{n+1} - \tau) + \mathbf{u}^{(4)}(\tau)\frac{(t_{n+1} - \tau)^2}{2} + \dots \end{aligned} \quad (2.38)$$

the integral in Eq. (2.37a) can be solved by a numerical quadrature as

$$\int_{t_n}^{t_{n+1}} \ddot{\mathbf{u}}(\tau) d\tau = (1 - \gamma)h\ddot{\mathbf{u}}_n + \gamma h\ddot{\mathbf{u}}_{n+1} + \underbrace{\frac{h^2}{2}(2\gamma - 1)\mathbf{u}^{(3)}(\tilde{\tau}) + \mathcal{O}(h^3\mathbf{u}^{(4)})}_{\text{error measure}} \quad (2.39)$$

where Eq. (2.38) has been multiplied by γ and $(1 - \gamma)$. Multiplication of the same equations by $(1 - 2\beta)$ and 2β yields the quadrature for the integral in Eq. (2.37b)

$$\int_{t_n}^{t_{n+1}} (t_{n+1} - \tau) \ddot{\mathbf{u}}(\tau) d\tau = \frac{h^2}{2}(1 - 2\beta)\ddot{\mathbf{u}}_n + \beta h^2\ddot{\mathbf{u}}_{n+1} + \underbrace{(\beta - \frac{1}{6})h^3\mathbf{u}^{(3)}(\tilde{\tau}) + \mathcal{O}(h^4\mathbf{u}^{(4)})}_{\text{error measure}}. \quad (2.40)$$

Therefore, the Newmark family schemes consist of the equations

$$\mathbf{u}_{n+1} = \mathbf{u}_n + h\dot{\mathbf{u}}_n + \frac{h^2}{2}(1 - 2\beta)\ddot{\mathbf{u}}_n + \beta h^2\ddot{\mathbf{u}}_{n+1}, \quad (2.41a)$$

$$\dot{\mathbf{u}}_{n+1} = \dot{\mathbf{u}}_n + (1 - \gamma)h\ddot{\mathbf{u}}_n + \gamma h\ddot{\mathbf{u}}_{n+1}, \quad (2.41b)$$

where γ and β are the algorithm parameters associated to the quadrature scheme, which give rise to well known temporal schemes such as explicit scheme or central differences (Alipour et al., 2000). Given an initial displacement and velocity, the initial acceleration can be calculated as the solution to

$$\mathbf{M}\ddot{\mathbf{u}}_0 = \mathbf{F}_0 - \mathbf{D}\dot{\mathbf{u}}_0 - \mathbf{K}\mathbf{u}_0 \quad (2.42)$$

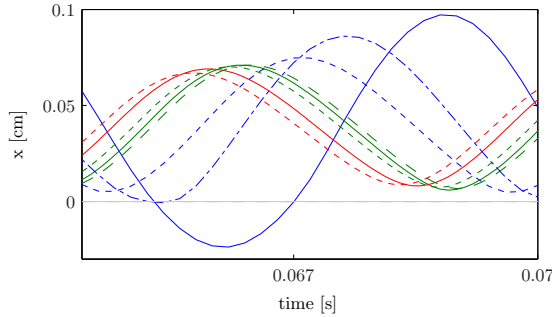


Figure 2.7: Blue $h = 200 \mu s$, green $h = 100 \mu s$, red $h = 50 \mu s$; solid 'TR', short dashed 'HHT' with $\alpha = -0.3$, long dashed 'HHT' with $\alpha = -0.05$, dot-dashed 'MAA' with $\gamma = 0.6$. After (Granados et al., 2014).

An alternative method is the Hilber-Hughes-Taylor α -method, which consists of a weighted mean of the Newmark approximations in Eqs. (2.41) as

$$\mathbf{u}_{n+1}^{HHT} = \mathbf{u}_{n+1} + \alpha (\mathbf{u}_{n+1} - \mathbf{u}_n) = (1 + \alpha) \mathbf{u}_{n+1} - \alpha \mathbf{u}_n, \quad (2.43a)$$

$$\dot{\mathbf{u}}_{n+1}^{HHT} = \dot{\mathbf{u}}_{n+1} + \alpha (\dot{\mathbf{u}}_{n+1} - \dot{\mathbf{u}}_n) = (1 + \alpha) \dot{\mathbf{u}}_{n+1} - \alpha \dot{\mathbf{u}}_n. \quad (2.43b)$$

When α is zero, the scheme is equivalent to the Newmark's family of methods. Then, the parameters h , γ , β and α control the numerical scheme behavior related to the error measures. In Fig. 2.7 the effect of different parameter choices on the trajectory of a node placed on the vocal fold surface is shown. The parameter values result into Newmark schemes equivalent to the trapezoidal rule ('R'), modified average acceleration ('MAA'), as well as the Hilber-Hughes-Taylor α -method ('HHT'). These results demonstrate that the numerical error has a noticeable effect on the model predictions (Paper A), which is translated into uncertainties in the prior knowledge of the system, crucial for inversion procedures (Paper D).

The error analysis can be typically found in the literature in terms of stability and accuracy. In this section a general overview of these error measures are presented, as well as the main results for a number of integration schemes. However, it is worthwhile to mention that the error analysis presented here assumes smoothness and homogeneity of the system, as well as small Rayleigh damping (modal damping or structural damping). These analysis disregard the effect of general viscous damping \mathbf{D} , and the non-smooth dynamics introduced by the contact phenomena and the air coupling. Hence, the effect of the temporal scheme on the behavior of the vocal fold problems remains to be studied by computer evaluation (Paper A).

Stability. The test equation $\dot{y} = ky$ with initial condition $y(0)$ with $k \in \mathbb{C}$, has the solution $y(t) = e^{kt}y(0)$. This function tends to 0 as $t \rightarrow \infty$ if $\Re(k) < 0$. Note that if this condition is not satisfied, a finite variation of the initial state vector will be amplified at subsequent instants. The concept of stability is related to the idea that solutions are bounded as $t \rightarrow \infty$. The concept of stability in integration methods is similar. As stated in (Gérardin, 1994), an time integration method is said to be stable if there exists an integration step $h_0 > 0$ so that for any $h \in [0, h_0]$ a finite variation of the state vector at time t_n induces only a non-increasing variation of the state vector y_{n+j} . If an integration can be formulated in the form $y_{n+1} = \mathbf{A}(hk)y_n$, and, by induction, $y_n = [\mathbf{A}(hk)]^n y_0$, where \mathbf{A} is an amplification matrix, an analytical stability analysis can be performed. Thus, points in the set $\{z = hk \in \mathbb{C}; |\mathbf{A}(z)| \leq 1\}$, define an integration scheme that does not amplify finite variations of the solution vector. The method is said to be stable in that region. If the region of stability contains the left half complex plane¹, then the method is called *A-stable* or *unconditionally stable* (Hughes, 2000).

To analyze the stability of the Newmark method, by means some system matrix manipulations it is possible to write the system in terms of an amplification matrix with variable $\mathbf{y}_n = [\mathbf{u}_n, \dot{\mathbf{u}}_n]^T$ (Hughes, 2000). The spectral radius of the amplification matrix is defined as

$$\rho(\mathbf{A}(h)) = \max_i |\lambda_i^h| \quad (2.44)$$

where λ_i^h denote the eigenvalues. To prevent a solution growth without bound, that is, instability, it is necessary that $\rho(\mathbf{A}(h)) \leq 1$. The conditional and unconditional stability regions are then defined by inequality relations in h , γ and β . The resulting spectral radius also control the decay of the solution modes, which may lead to a beneficial damping of spurious high-frequency characteristic of the finite element solution (Hughes, 2000).

Accuracy. A local truncation error $\tau(t_n)$ arises from the temporal discretization, and may be seen as a measure of accuracy of the algorithm. The error is the difference between the exact and the approximated solution, and takes the form $h \cdot \tau(t_n)$. For $|\tau(t)| = \mathcal{O}(h^k)$, the parameter $k > 0$ is called the order of accuracy or rate of convergence. The Newmark method satisfies (Hughes, 2000)

$$\tau(t) = \begin{cases} \mathcal{O}(h^2) & \text{if } \gamma = 1/2 \\ \mathcal{O}(h) & \text{otherwise} \end{cases} \quad (2.45)$$

Furthermore, numerical dissipation (amplitude error), and numerical dispersion (periodicity error) allow further accuracy analysis. Recall that the exact

¹Note that if $\Im(hk) \leq 0$, since $h > 0$, then $\Re(k) \leq 0$.

natural frequency of the problem differs from that one obtain from finite element spatial discretization, and in turn from the one produced by the temporal discretization. By considering a damped test problem without source term, the difference between the amplitude of the exact solution and the numerical one is the numerical dissipation or damping introduced by the temporal scheme. Similarly, the dispersion or periodicity error related the actual frequencies and the ones obtained after numerical discretization.

2.4 Numerical solution

Once the model is discretized, specific predictions in the form of numerical solutions must be obtained. As mentioned in the previous section, the discretized problem in Eq. (2.33) is derived from the weak form in Eq. (2.31). Notice that only the Neumann boundary condition describing the airflow coupling in Eq. (2.25) appears in the weak formulation. Hence, the Dirichlet and contact constraints may be enforced by means of additional techniques, and then solutions to the problem can be found. In this section, some of these techniques are shortly described.

In analytic mechanics, the equilibrium equations emanate from the Hamilton's principle, that is, through the process of minimizing a functional containing both kinetic and potential energy. Under a static linear elasticity regime, the equilibrium deformation \mathbf{u} can be found as the stationary or critical point of the energy functional

$$\Pi^*(\mathbf{v}) = \frac{1}{2} \int_{v_{solid}} \boldsymbol{\epsilon}(\mathbf{v})^T \cdot \mathcal{S} \cdot \boldsymbol{\epsilon}(\mathbf{v}) \, dv - \int_{\Gamma_N} \mathbf{v}^T \cdot \mathbf{p} \, d\Gamma, \quad (2.46)$$

where the aerodynamic forces have been assumed conservative. Taking the variation of the functional, the equilibrium deformation is such that

$$\delta\Pi^*(\mathbf{u}) = \int_{v_{solid}} \boldsymbol{\epsilon}(\delta\mathbf{u})^T \cdot \mathcal{S} \cdot \boldsymbol{\epsilon}(\mathbf{u}) \, dV - \int_{\Gamma_N} \delta\mathbf{u}^T \cdot \mathbf{p} \, dS = 0 \quad (2.47)$$

and this stationary point will be a minimum because the system matrix is positive definite under these assumptions (the strain energy is nonnegative). If the admissible test function \mathbf{w} in Eq. (2.31) are understood as variations $\delta\mathbf{v}$ of the displacements \mathbf{v} in Eq. (2.46), then the Galerkin formulation in Eq. (2.32) is equivalent to the first variation of the energy functional. Considering the Dirichlet and contact constraints in Eqs. (2.24)-(2.26), the deformation is the solution of a minimization problem that has objective function the functional

in Eq. (2.46), that is,

$$\begin{aligned} & \text{minimize} && f(x) \equiv \Pi^*(\mathbf{v}) \\ & \text{subject to} && h(x) \equiv \mathbf{v} = 0 \quad \text{on } \Gamma_D \\ & && g(x) \equiv (\mathbf{v} - \bar{\mathbf{v}}) \cdot \bar{\mathbf{n}} - (\mathbf{X} - \bar{\mathbf{X}}) \cdot \bar{\mathbf{n}} \geq 0 \quad \text{on } \Gamma_C. \end{aligned} \quad (2.48)$$

Note that the inequality condition is equivalent to Eq. (2.26), and that the Neumann boundary conditions are not enforced because they are implicit in the first variation $\delta\Pi^*$ equivalent to the weak formulation. Using variations and optimization methods for equality and inequality constraints enforcement, solutions to the equations of motion can be found. These methods typically consist of formulating and solving an unconstrained minimization problem where the new objective function contains terms that account for the constraints. From a physical point of view, the new energy functional contains energy terms that account for the energy that arises at the constrained boundary, that is, reaction forces. A short review of a number of optimization approaches is presented in what follows. Further details can be found in (e.g., Bertsekas, 1982). The Dirichlet constraint is here neglected for the sake of simplicity, as equality constraints are a particular case of inequality constraints.

Whenever the inequality is not fulfilled, the unconstrained minimization problem to be solved with the penalty method is

$$\text{minimize} \quad f(x) + \kappa|g(x)|^2 \quad (2.49)$$

where κ is a user-defined constant. The new objective function is smooth and does not have additional degrees of freedom. For convex functions, stationary points would correspond to a local minimum.

The classical Lagrange multipliers method is based on the Lagrangian function

$$\mathcal{L}(x, \lambda) = f(x) + \lambda g(x) \quad (2.50)$$

where λ is called Lagrange multiplier or dual variable. Then, the solution (x, λ) is found as the solution to

$$\begin{aligned} & \text{maximize}_{\lambda} \inf_x \mathcal{L}(x, \lambda) \\ & \text{subject to } \lambda \leq 0 \end{aligned} \quad (2.51)$$

The function $d(\lambda) = \inf_x \mathcal{L}(x, \lambda)$ is concave and it is called the dual functional. If the original problem is smooth and convex, then the infimum occurs where

the gradient is equal to zero, and a local optima (x, λ) satisfies the conditions

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda) &= f'(x) + \lambda g'(x) = 0 \\ g(x) &\geq 0 && \text{primal feasibility} \\ \lambda &\leq 0 && \text{dual feasibility} \\ g(x)\lambda &= 0 && \text{complementary slackness} \end{aligned} \tag{2.52}$$

which are known as the Karush-Kuhn-Tucker (KKT) conditions for optimality. Assuming $f(x)$ represents the energy of a mechanical system with contact, then the term $\lambda g'(x)$ can be seen as the reaction force resulting of the contact condition enforcement. Hence, the KKT conditions physically model a non-penetration between bodies ($g(x) \geq 0$), compressive forces pushing the bodies apart at contact ($\lambda \leq 0$), and whenever there is no contact, then no reaction forces occur ($g(x)\lambda = 0$).

However, the Lagrange method requires local convexity and smoothness, which is not satisfied in general physical problems. Alternatively, an augmented Lagrange method solves the minimization problem of a perturbed functional as

$$\begin{aligned} &\text{minimize } f(x) + \kappa |g(x)|^2 \\ &\text{subject to } g(x) \geq 0 \end{aligned} \tag{2.53}$$

It can be shown that for large κ local convexity can be assured. An augmented Lagrangian can be constructed as $\mathcal{L}^a(x, \lambda) = f(x) + \kappa |g(x)|^2 + \lambda g(x)$, and the problem in Eq. (2.51) can be formulated. Then, a steepest ascent can be applied to maximize the dual function and find a λ estimate, thereby avoiding solving for the exact multipliers in the KKT conditions.

In the particular case of the vocal fold deformation, the corresponding energy functional $\Pi(\mathbf{v})$ contains viscous (non-conservative) and dynamic terms. Assuming that the variation of the energy is equivalent to the Galerkin form in Eq. (2.32), the problem could be solved by adapting the optimization techniques to the present problem. However, the new energy functional is not necessarily convex. Hence, the quality of the solution must be evaluated computationally to investigate the accuracy of the predictions (Paper A, Paper B).

The process of finding a solution by the application of the optimization methods described above may lead to a linear problem

$$Az = f, \tag{2.54}$$

where $z = x$ in the penalty and augmented Lagrangian solutions, and $z = (x, \lambda)$ in the Lagrange solution (KKT conditions). The difficulty of finding a solution

relies on the numerical nature of the matrix. In the case of linear elasticity, the system matrix is positive definite whenever $z = x$. Therefore, the optimal point occurs at a minimum, as mentioned earlier in this section, and iterative gradient-based methods can be used to solve the linear system. However, in the case of the KKT conditions, the solution $z = (x, \lambda)$ occurs at a saddle point; see Eq. (2.51). This means that the system matrix is indefinite and gradient-based method would fail at finding the solution. Alternative methods, such as the iterative generalized minimal residual method may be more efficient (Paper A).

2.5 Contributions

Previous studies on continuum models of the vocal folds include collision. In (Gunter, 2003) symmetric vocal fold contact is assumed and the impact forces are orthogonal to a virtual rigid mid-plane that prevents node interpenetration, similarly to (Alipour et al., 2000). The model does not included airflow coupling, and vocal fold separation is not predicted by the model. A symmetric Herzian model is used in (Spencer et al., 2008) to estimate the contact pressure from measured strain fields, although models of the vocal fold motion are out of scope. The coupled fluid-structure model in (Zheng et al., 2010) assumes symmetric contact, and the contact forces are modeled as springs, similarly to (Tao et al., 2006), where the augmented Lagrangian algorithm is applied. However, no specific reference to the contact model parameters nor the contact discretization elements is made. The symmetric fluid-structure model in (Xue et al., 2012) describes the contact interface as node-to-node elements following a predefined non- linear, spring-based law. Two artificial non-slip and non-penetrable collision planes are placed at the mid-plane to detect contact, and a minimum glottal gap is enforced to avoid failure of the flow solver. A more general contact model based on a finite element contact discretization is developed in (Rosa et al., 2003). The contact model could potentially allow for asymmetric contact and provide accurate predictions of the impact forces. However, results in this direction are not provided. Finally, in regard to the temporal discretization schemes, central differences (Alipour et al., 2000) or Newmark's schemes (Rosa et al., 2003; Zheng et al., 2010), are typically used although the effect on the model predictions has not been investigated.

The scientific papers published in reference to the continuum model of the vocal folds have as a main focus the accuracy of the model's mechanical and dynamical predictions, specially when vocal fold collision occurs during symmetric and asymmetric vibration. An improvement of the predictions, an subsequently a decrease of the uncertainties, is important for inversion procedures. In addition, the vocal fold contact models may be used as a complementary mathematical

tool to support clinical studies on the relation between vocal fold lesions and phonatory trauma.

The contribution of the papers can be summarized as follows.

Paper A

- A three-dimensional viscoelastic continuum model of the vocal folds with contact is provided, along with validation.
- A flow model developed under a one-dimensional Bernoulli regime is capable to set the vocal folds into oscillation and to induce vocal fold separation after contact, thereby reducing the computational complexity.
- A novel contact detection mechanism is proposed, which is shown to successfully detect asymmetric contact.
- A solution algorithm independent of contact model parameters (Lagrange) is investigated. It is shown to be advantageous over widely used linear spring-like contact models (penalty) that have been shown to perform poorly at modeling fast vocal fold impacts.
- Model predictions are shown to benefit from highly diffusive temporal schemes. These schemes allow for larger time steps, which may improve the computational time in inverse problems simulations.
- The contact model is able to predict specific mechanical behavior.
- The proposed contact discretization method constitutes a suitable framework for frictional contact.

Paper B

- In an effort to regularize local contact force estimates of the contact model presented in Paper A, an augmented Lagrangian solution method is adapted to the variational inequality constrained contact problem.
- The method is shown to produce a smoothing effect, although it shows a strong dependency on the parameter choice.

CHAPTER 3

High-speed videoendoscopic data acquisition

Non-invasive procedures for quantitative analysis of laryngeal endoscopic data are necessary to extract the underlying vocal fold dynamics. The vocal fold oscillation can be analyzed by applying image processing algorithms for vocal fold edge tracking to laryngeal videoendoscopy (Wittenberg et al., 1995; Larsson et al., 2000; Lohscheller et al., 2007). However, these algorithms do not capture the entire deformation visible in the images, which provides valuable data for the investigation of vocal fold tissue characteristics (Spencer et al., 2008). To solve the problem of continuum vocal fold features inference from laryngeal high-speed videoendoscopy, a large number of observations is required to reduce the uncertainty related to the underdetermined system. The objective of this chapter is to introduce mathematically the concept of optical flow to be used for vocal fold motion extraction (Paper C). The discussion is taken from a deterministic point of view, and sources of numerical error are discussed. The theory presented here is a review of the optical flow literature mainly based on the landmark papers (Horn and Schunck, 1981; Lucas and Kanade, 1981; Anandan, 1989; Black and Anandan, 1996). Furthermore, the general notion of principal components decomposition as a data analysis technique to extract the underlying vibration pattern is presented. A summary of the scientific contributions of the thesis is presented at the end of the chapter in relation to a brief summary of the state of the art. Further details can be found in the

corresponding paper (Paper C).

3.1 Optical flow

Visual inspection of moving objects can provide information on the motion in different ways. For example, an observer of a track and field sporting event can distinguish which runner is faster, can analyze how the runners reach the finish line, or can keep track of one specific runner during the race. In that sense, the moving runners behave as a flow, where the velocity at one specific time changes between runners, runners cross the finish line at different speeds, and also they follow trajectories. Then, the observer perceives the moving scene as an optical flow. Filming a track and field event would then result in an image sequence containing information related to the optical flow.

3.1.1 Continuity equation

Optical flow is defined as the estimate of the velocity field of moving objects captured in an image sequence by means of a scalar intensity field $I(\mathbf{x}, t) \in \mathbb{Z}^+$. A pixel located at the point $\mathbf{x} = (x_1, x_2)$ at time t with brightness $I(\mathbf{x}, t)$ and unknown velocity (optical flow) $\mathbf{v} = (v_1, v_2)$ will at time $t + \delta t$ have brightness $I(\mathbf{x} + \mathbf{v}\delta t, t + \delta t)$. The change in intensity is then

$$\begin{aligned} \delta I &= I(\mathbf{x} + \mathbf{v}\delta t, t + \delta t) - I(\mathbf{x}, t) = \frac{\partial I}{\partial x_1} v_1 \delta t + \frac{\partial I}{\partial x_2} v_2 \delta t + \frac{\partial I}{\partial t} \delta t \\ &= (\nabla I(\mathbf{x}, t) \cdot \mathbf{v} + I_t(\mathbf{x}, t)) \delta t, \end{aligned} \quad (3.1)$$

where second order terms in δt of the Taylor expansion are neglected. The main assumption of the optical flow theory is that brightness remains constant along pixel trajectories. That is,

$$\delta I = 0 \Rightarrow \frac{DI}{Dt} = \mathbf{v} \cdot \nabla I(\mathbf{x}, t) + I_t(\mathbf{x}, t) = 0, \quad (3.2)$$

which corresponds to the continuity equation and assumes smooth and differentiable pixel trajectories.

3.1.2 Regularization

By simple algebraic manipulation it can be shown that the continuity equation in Eq. (3.2) yields the projection of the velocity onto a line in the gradient

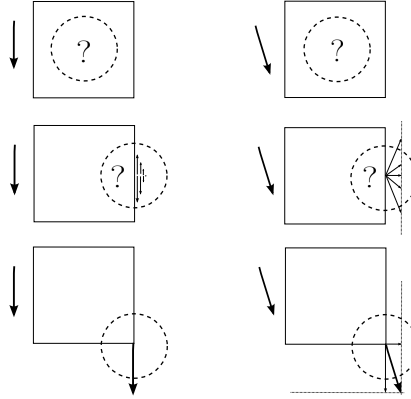


Figure 3.1: Illustration of the undetermined continuity equation for optical flow estimation. The dashed circle bounds the observation area where constant velocity is assumed in order to solve the problem.

direction as

$$\frac{-I_t}{\|\nabla I(\mathbf{x}, t)\|} \frac{\nabla I(\mathbf{x}, t)}{\|\nabla I(\mathbf{x}, t)\|}. \quad (3.3)$$

Because the continuity equation is underdetermined in $\mathbf{v} = (v_1, v_2)$, the tangent component remains unknown. This is illustrated in Fig. 3.1. The situation shown at the left column illustrates a rigid block with a uniform intensity field moving downwards, while at the right column the block follows an oblique movement. Imagine that an observer can only see what happens within a limited observation region \mathcal{R} , bounded by the dashed circle in the figure, and wants to infer the direction and magnitude of the velocity of the moving block. In the top row, the spatial gradient of the intensity field is zero within the observation region, because uniform intensity is assumed. Then, from Eq. (3.2), the velocity cannot be inferred. If the observation region is, however, as shown in the second row in the figure, the spatial gradients in the pixels placed at the edge of the block are non-zero, although they are parallel. Then, from Eqs. (3.2)-(3.3) the normal component can be estimated. However, the actual direction and magnitude of the velocity is uncertain because the tangent component remains unknown. In the situation shown in the last row, two different intensity gradients can be calculated, and two normal components (onto a line in the gradient direction) can be estimated. Hence, the actual block velocity can be inferred.

The situation described above can be formulated mathematically for a general object in motion. A constant velocity field is assumed to be constant within the observation region $\mathcal{R} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$, and a number of constraints are

enforced. Then, the velocity is the solution to the least-squares problem

$$\underset{\mathbf{v}}{\text{minimize}} \sum_{\mathbf{x} \in \mathcal{R}} |\nabla I(\mathbf{x}, t) \cdot \mathbf{v} + I_t|^2 = \underset{\mathbf{v}}{\text{minimize}} \|\mathbf{A}\mathbf{v} - \mathbf{b}\|_2^2, \quad (3.4)$$

with

$$\mathbf{A} = \begin{pmatrix} I_{x_1}(\mathbf{x}_1, t) & I_{x_2}(\mathbf{x}_1, t) \\ I_{x_2}(\mathbf{x}_1, t) & I_{x_2}(\mathbf{x}_2, t) \\ \vdots & \\ I_{x_1}(\mathbf{x}_M, t) & I_{x_2}(\mathbf{x}_M, t) \end{pmatrix} \quad \text{and} \quad \mathbf{b} = - \begin{pmatrix} I_t(\mathbf{x}_1, t) \\ I_t(\mathbf{x}_2, t) \\ \vdots \\ I_t(\mathbf{x}_M, t) \end{pmatrix}, \quad (3.5)$$

and $\|\cdot\|_2$ is the discrete l^2 -norm. Situations where intensity gradients are parallel within the observation region lead to an underdetermined system; this is known in image processing as the aperture problem (Fleet and Weiss, 2006). When more than two intensity gradients are not parallel, an overdetermined system is obtained.

Assume that \mathbf{e} is the error of the numerically approximation of the time derivatives in \mathbf{b} , that is, $\tilde{\mathbf{b}} = \mathbf{b} + \mathbf{e}$. Equivalently, the error can be expressed as $\mathbf{e} = \mathbf{A}\tilde{\mathbf{v}} - \mathbf{b}$, where $\tilde{\mathbf{v}}$ is the numerical solution with a noise-contaminated numerical approximations. The least-square velocity in Eq. (3.4) is given by the solution to the normal equations (Hansen, 1998)

$$\tilde{\mathbf{v}} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \tilde{\mathbf{b}}, \quad (3.6)$$

where H indicates conjugate transpose. Furthermore, the singular value decomposition of the system matrix is

$$\mathbf{A} = \mathcal{U} \mathcal{G} \mathcal{V}^H, \quad (3.7)$$

which consists of two orthogonal matrices $\mathcal{U} \in \mathbb{R}^{M \times M}$ and $\mathcal{V} \in \mathbb{R}^{2 \times M}$ and a diagonal matrix $\mathcal{G} \in \mathbb{R}^{M \times M}$ with decreasing singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_M$. Then, it can be seen that the least-squares solution in Eq. (3.6) can be expressed as the summation of terms inversely proportional to the decreasing singular values as (Hansen, 1998)

$$\tilde{\mathbf{v}} = \mathcal{V} \mathcal{G}^{-1} \mathcal{U}^H \tilde{\mathbf{b}} = \sum_{i=1}^M \frac{(\mathbf{w}_i^1)^H \tilde{\mathbf{b}}}{\sigma_i} \mathbf{w}_i^2 \quad (3.8)$$

where \mathbf{w}_i^1 and \mathbf{w}_i^2 are the column vectors of \mathcal{U} and \mathcal{V} respectively. Hence, the small errors in \mathbf{b} would be amplified if the system matrix contains small singular values, leading to a poor least-squares estimate. The smallness of the singular values is measured by the condition number of the system matrix,

defined as the ratio between the largest and the smallest singular values, and provides an upper bound of the solution error. Large condition numbers lead to ill-conditioned system matrices. In the optical flow estimation, ill-conditioned matrices occur when intensity gradients are almost parallel within the observation region \mathcal{R} . Then, the solution is very sensitive to the noise in the image and to the numerical approximations. To reduce the noise effect and to obtain numerical solutions close to the exact one, regularization methods can be applied (Hansen, 1998). These numerical methods consist of incorporating further constraints on the solution to control the quality of the solution. For example, the growing norm of the noise-contaminated optical flow solution is bounded by an arbitrary constant ε in the classical Tikhonov regularization method as

$$\begin{aligned} & \underset{\tilde{\mathbf{v}}}{\text{minimize}} \quad \|\mathbf{A}\tilde{\mathbf{v}} - \tilde{\mathbf{b}}\|_2^2, \\ & \text{subject to} \quad \|\tilde{\mathbf{v}}\|_2^2 \leq \varepsilon. \end{aligned} \quad (3.9)$$

If the Lagrange multiplier method is applied, as seen in the previous chapter, the constrained minimization problem is equivalent to

$$\underset{\tilde{\mathbf{v}}}{\text{minimize}} \quad \|\mathbf{A}\tilde{\mathbf{v}} - \mathbf{b}\|_2^2 + \lambda \|\tilde{\mathbf{v}}\|_2^2 \quad (3.10)$$

with $\lambda \geq 0$ a Lagrange multiplier. However, the Tikhonov method does not find an optimal λ in a maximization algorithm, given the arbitrariness of the bound ε in relation to the unknown noise. Instead, method for choosing the parameter λ , known as parameter-choice methods in the inverse problems literature (Hansen, 1998), can be used, and the parameter is then called the regularization parameter. In most of these methods, as well as in the Tikhonov regularization method, it is assumed that the error follows a Gaussian distribution $\mathbf{e} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, with an unknown variance proportional to the regularization parameter; this aspect will be discussed in the next chapter. It is simple to see that the Tikhonov solution to the problem in Eq. (3.10) is

$$\tilde{\mathbf{v}} = (\mathbf{A}^H \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^H \tilde{\mathbf{b}}, \quad (3.11)$$

where \mathbf{I} is the identity matrix. For large values of λ , the solution is highly regularized and the error in \mathbf{b} is assumed to be large. If $\lambda = 0$, the solution corresponds to the non-regularized least-squares solution in Eq. (3.6). Hence, a trial-and-error technique can also be applied to regularize the solutions, as an alternative to parameter-choice techniques.

The numerical treatment of the optical flow solution described above is based on the formulation in Eq. (3.4), which assumes a predefined region \mathcal{R} of constant velocity; recall the rigid block example in Fig. 3.1. However, rigid motion is not always a good assumption, such as in the complex vocal fold tissue deformation (Spencer et al., 2008). In that situation, the velocity field may be different

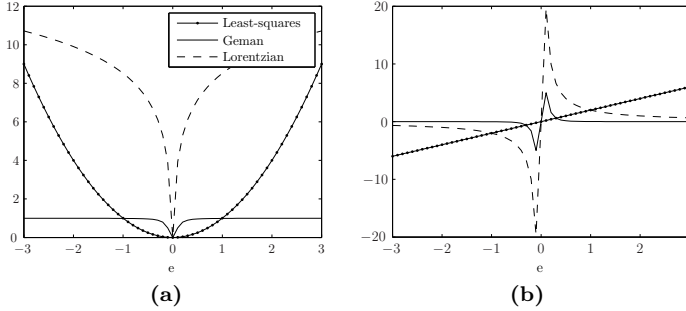


Figure 3.2: Geman&McClure and Lorentzian functions for robust estimation, as well as the l^2 -norm for classical least-squares problems in (a), and their derivatives in (b).

at each pixel position, and a dense optical flow estimate is necessary (Horn and Schunck, 1981, Paper C). Furthermore, ill-conditioning of the system matrix may still occur due to the underdeterminess of the continuity equation in Eq. (3.2) and the computation of almost parallel intensity gradients in certain regions. Hence, regularization techniques must be applied (Paper C).

3.1.3 Robust statistics

The continuity equation in Eq. (3.2) assumes continuous and smooth pixel trajectories. However, the assumption is repeatedly violated in complex scenarios, where occlusion, motion discontinuities or spatial discontinuities can occur, such as in laryngeal videoendoscopy (Paper C). Therefore, the assumption of normal distribution of the error may be weak. As a solution, different functions rather than the discrete l^2 -norm can be investigated, which is the object of study of robust statistics (Hampel et al., 1986; Anandan, 1989). These alternative functions may preferably give lower weights to the pixel positions where the computational error is very high, that is, where the continuity and smoothness assumptions are violated. In Fig. 3.2a three functions are shown: the Geman&McClure and the Lorentzian functions corresponding to

$$\rho(e, \tau) = \frac{e^2}{\tau + e^2} \quad \text{and} \quad \rho(e, \tau) = \log \left(1 + \frac{1}{2} \left(\frac{e}{\tau} \right)^2 \right), \quad (3.12)$$

respectively, where τ is a scale parameter that controls the convexity of the functions. In the figure, the least-squares l^2 -norm is shown for comparative reasons. The derivatives of the functions are called the influence function and characterize the bias that a particular measurement has on the solution. These

derivatives are shown in Fig. 3.2b. Note that in the least-squares case, $\phi(e) = 2e$, which indicates that the influence increases without bound.

3.1.4 Aliasing

When dealing with time-dependent intensity field representation of moving objects, temporal and spatial aliasing may have an effect on the optical flow estimates (Paper C). Sequence of movie frames is the result of temporal sampling of the continuous motion of a real object. Sampling an analog signal creates replicas of the signal spectrum (Proakis and Manolakis, 1996). It can be easily seen that these replicas together with high velocity fields reduce the number of spatial frequencies free of aliasing (Simoncelli, 1993), creating a potential loss of spatial spectral information; hence the optical flow algorithm may provide poor estimates of the velocity.

A method to solve aliasing problem caused by temporal sampling is the coarse-to-fine approach (Anandan, 1989). In this method a spatial filter, such as a Gaussian filters, at a coarse scale can be applied to the image at a certain time step, and thereby remove high spatial frequency content and reduce the velocity estimate (Simoncelli, 1993). A filtered image at a finer scale, with higher spatial frequency components, can then be warped according to the estimated velocity at the previous coarse scale. Then, most of the motion is reversed and residual small motions free of aliasing are left to be estimated in an iterative process. Mathematically, for a velocity field estimate at a coarser scale \mathbf{v}_0 , a warped image at a finer scale is defined as

$$I_1(\mathbf{x}, t + \delta t) = I(\mathbf{x} + \mathbf{v}_0 \delta t, t + \delta t) \quad (3.13)$$

Assuming that the actual flow velocity is $\mathbf{v} = \mathbf{v}_0 + \delta \mathbf{v}$, then

$$I_1(\mathbf{x} + \delta \mathbf{v} \delta t, t + \delta t) = I(\mathbf{x} + (\mathbf{v}^0 + \delta \mathbf{v}) \delta t, t + \delta t) \quad (3.14)$$

Similarly to the minimization in Eq. (3.4), it can be seen that the residual flow $\delta \mathbf{v}$ satisfies (Fleet and Weiss, 2006)

$$\min_{\delta \mathbf{v}} \sum_{\mathbf{x} \in \mathcal{R}} |\nabla I_1(\mathbf{x}, t) \cdot \delta \mathbf{v} + I_{0,t}|^2. \quad (3.15)$$

Hence, image spatial derivatives can be computed using the warped image at a finer scale. The new velocity estimate

$$\mathbf{v}_1 = \mathbf{v}_0 + \delta \mathbf{v} \quad (3.16)$$

is then free of aliasing. A spatial filter can again be applied at a finer scale, that is, a image with more high-frequency content. The process is repeated by creating new warped images from spatial filtering with

$$I_j(\mathbf{x}, t + \delta t) = I(\mathbf{x} + \mathbf{v}_j \delta t, t + \delta t) \quad (3.17)$$

and solving Eq. (3.15) with $j > 1$.

3.2 Principal components analysis

Principal component decomposition is a method for dimensionality reduction which can be used for statistical data analysis (Jackson, 1991; Cootes et al., 1995). Principal component analysis (PCA) deals with redundancy in multivariate data and allows a simpler understanding and visualization of the data, as well as identification of hidden patterns in a data set (Berry et al., 2001). The result is a change of basis or principal components that maximizes the variance of the data. There are many derivations of the principal components. In this thesis, an optimization-based derivation is presented in what follows.

Let \mathbf{v}^i , $i = 1, \dots, N$ be the realizations of an n -dimensional random variable. The aim of PCA is to find a new basis such that when the data is projected onto this space, its variance represents a large portion of the total variance of the original data. Mathematically, the data is transformed according to

$$\tilde{\mathbf{v}} = \mathbf{M}\mathbf{v} \quad (3.18)$$

with $\mathbf{M}^{-1} = \mathbf{M}^T \in \mathbb{R}^{n \times n}$ (orthogonal). It is straightforward to see that the sample covariance matrices of \mathbf{v} and the transformed data are

$$\Sigma_{\mathbf{v}} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{v}^i - \bar{\mathbf{v}})(\mathbf{v}^i - \bar{\mathbf{v}})^T \in \mathbb{R}^{n \times n} \quad (3.19)$$

and

$$\Sigma_{\tilde{\mathbf{v}}} = \frac{1}{N-1} \sum_{i=1}^N (\tilde{\mathbf{v}}^i - \tilde{\bar{\mathbf{v}}})(\tilde{\mathbf{v}}^i - \tilde{\bar{\mathbf{v}}})^T = \mathbf{M}\Sigma_{\mathbf{v}}\mathbf{M}^T \in \mathbb{R}^{n \times n}, \quad (3.20)$$

with means

$$\bar{\mathbf{v}} = \frac{1}{N} \sum_{i=1}^N \mathbf{v}^i \quad \text{and} \quad \tilde{\bar{\mathbf{v}}} = \frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{v}}^i, \quad (3.21)$$

respectively. The goal of principal components decomposition is then to find a basis $\{\mathbf{v}_{pc}^i\}_{i=1}^n$ that maximizes the variance of the data set. Hence, the first

basis vector $\mathbf{v}_{pc}^1 \in \mathbb{R}^n$ is the solution to the problem

$$\begin{aligned} & \underset{\mathbf{v}_{pc}^1}{\text{maximize}} \quad (\mathbf{v}_{pc}^1)^T \boldsymbol{\Sigma}_{\mathbf{v}} \mathbf{v}_{pc}^1 \\ & \text{subject to} \quad (\mathbf{v}_{pc}^1)^T \mathbf{v}_{pc}^1 = 1, \end{aligned} \quad (3.22)$$

where the constraint is added to avoid unbounded solutions. By means of the Lagrange multiplier method, the solution to the problem satisfies

$$\boldsymbol{\Sigma}_{\mathbf{v}} \mathbf{v}_{pc}^1 = \lambda_1 \mathbf{v}_{pc}^1, \quad (3.23)$$

It is clear that the solution is the eigenvector of $\boldsymbol{\Sigma}_{\mathbf{v}}$ with the largest eigenvalue. By repeating the process with the rest of the basis vectors and assuming orthogonality, the transformation satisfies

$$\boldsymbol{\Sigma}_{\tilde{\mathbf{v}}} = \mathbf{M}^T \boldsymbol{\Sigma}_{\mathbf{v}} \mathbf{M}, \quad (3.24)$$

and it is the eigenvalue decomposition of $\boldsymbol{\Sigma}_{\mathbf{v}}$. This means that the transformation matrix \mathbf{M} has the n eigenvectors of $\boldsymbol{\Sigma}_{\mathbf{v}}$ as rows, corresponding to the basis vectors $\{\mathbf{v}_{pc}^i\}_{i=1}^n$, and the covariance matrix of the transformed data is a diagonal matrix with the variances being the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ of $\boldsymbol{\Sigma}_{\mathbf{v}}$. Each of these n eigenvectors are the principal components.

In many applications, such as in image analysis where n is typically related to the image resolution, it is common to deal with large values of the dimensionality n , and the numerical implementation of the covariance matrices may be somewhat cumbersome. In general, because most of the data variation is expected to be explained by a reduced number of principal components $r \leq n$, the implementation of a matrix $\mathbf{M} \in \mathbb{R}^{r \times n}$ with $r < n$ may be sufficient. However, the principal components cannot be found by eigenvalue factorization, as the matrix is not square. Instead, singular value decomposition as in Eq. (3.7) may be performed (Jackson, 1991). Let $\mathbf{V} \in \mathbb{R}^{N \times n}$ with $N < n$ be a matrix containing all the data samples. Then, the covariance matrix in Eq. (3.19) satisfies

$$\boldsymbol{\Sigma}_{\mathbf{v}} = \mathbf{V}^T \mathbf{V} \in \mathbb{R}^{n \times n} \quad (3.25)$$

It can be shown that the first N eigenvalues of $\mathbf{V}^T \mathbf{V}$ are non-zero, and the remaining $n - N$ eigenvalues are zero. The singular value decomposition of \mathbf{V} consists on two orthogonal matrices $\mathcal{U} \in \mathbb{R}^{N \times N}$ and $\mathcal{V} \in \mathbb{R}^{n \times N}$ and a diagonal matrix $\mathcal{G} \in \mathbb{R}^{N \times N}$ with decreasing singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N$ as in Eq. (3.7). Then the covariance matrix of the transformed data satisfies

$$\boldsymbol{\Sigma}_{\tilde{\mathbf{v}}} = \mathbf{M} \mathcal{U} \mathcal{G}^2 \mathcal{V}^T \mathbf{M}^T \quad (3.26)$$

If $r = N < n$ and $\mathbf{M}^T \in \mathbb{R}^{n \times N}$ is composed by the N column vectors of \mathcal{V} , that is,

$$\mathbf{M}^T = \mathcal{V}, \quad (3.27)$$

then the covariance matrix of the transformed data in Eq. (3.26) is diagonal, and the variances are the squared root of the singular values. The new principal component basis $\{\mathbf{v}_{pc}^i\}_{i=1}^N$ with $\mathbf{v}_{pc}^i \in \mathbb{R}^n$ also maximizes the variance of the data and the matrix manipulations to obtain this change of basis is simplified. This new basis captures the essence of a large data set, and may reveal hidden patterns of, for instance, an optical flow image sequence (Paper C).

3.3 Contributions

The use of non-invasive optical techniques for laryngeal high-speed videoendoscopic analysis is typically based on the free vocal fold superior edge motion (Wittenberg et al., 1995; Lohscheller et al., 2007). However, this data is not sufficient to capture the entire two-dimensional vibration pattern, needed for continuum model inversion. Works by (Berry et al., 2006; Spencer et al., 2008; Bakhshaei, 2013) apply digital image correlation techniques to high-speed videos of synthetic and excised larynges to extract full superior surface strains. This technique, however, requires a speckle pattern to be deposited on the tissue previous to recording (Sutton et al., 1986), which make the procedure infeasible for direct laryngoscopy.

The scientific paper published in regard to high-speed videoendoscopic data acquisition focuses on a non-invasive procedure for dense data acquisition, and methods for error reduction in connection to inversion procedures. Furthermore, a quantitative analysis tool based on principal components decomposition relevant for clinical or scientific studies is investigated.

The contributions of the thesis in this direction can be summarized as follows.

Paper C

- A dense optical flow algorithm is presented and adapted to the particular nature of laryngeal videoendoscopic data. The algorithm has been shown sufficient to describe full two-dimensional dynamics of the vocal fold in a non-invasive way.
- Motion discontinuities, temporal aliasing and noise presence are treated numerically by means of regularization, robust statistics and anti-aliasing techniques.
- Principal components analysis is performed on the optical flow image sequence, the results showing that the method is able to identify hidden patterns of vibration.

CHAPTER 4

Bayesian statistical inversion

An inverse problem arises when solving any system of equations. For instance, solving for the velocity field in the optical flow continuity equation in Eq. (3.2) is an inverse problem. In complex problems, the solution typically requires specific numerical treatment, as discussed in the previous chapter. If the problem is underdetermined, infinitely many solutions are feasible, and typically additional constraints are enforced leading to the formulation of a least-squares problem; see Eq. (3.4). However, in the presence of modeling or measuring errors, ill-conditioning of the system matrix may lead to solutions growing without bound. Hence, deterministic regularization techniques can be applied, as in Eq. (3.10). Finally, one seeks to find a solution to the minimization problem, preferably the global minimum. In many applications, as in robust optical flow estimation, convexity cannot be assumed, and direct minimization procedures cannot guarantee a solution corresponding to the global minimum.

Similarly to the optical flow problem, vocal fold features inference can be formulated in terms of an inverse problem, involving measurement data and a vocal fold model. Given the complexity of the problem, underdeterminess, ill-conditioning or non-convexity of the minimization problem are numerical difficulties that can easily arise and must be properly treated. Many different direct minimization procedures have been used to solve the non-convex optimization problem (Döllinger et al., 2002; Yang et al., 2011; Pinheiro et al., 2012), al-

though an optimal solutions still remains to be found. The effect of possible underdeterminess of the system has not been studied in the literature, and it is expected to play a role given the complexity of the vocal fold structure and the low-dimensionality of the observation data (Hadwin et al., 2016). In addition, vocal fold modeling and data acquisition carry a large unknown error, which may have a direct effect on the estimation of vocal fold features (Schwarz et al., 2008). However, and specially for clinical applications, it is important to provide uncertainty information to the obtained results (Risholm et al., 2013), and thereby reflect the actual uncertainty of observation-based knowledge. Statistical inversion arises in an effort to account for modeling and numerical uncertainties, formulating the variables as stochastic variables, and assuming that the solutions are just realizations of the stochastic unknown variable (Kaipio and Somersalo, 2005). The solution is then expressed as a probability density function, and the estimation uncertainty can be characterized.

This chapter focuses on general Bayesian statistical inversion theory. Basic concepts of statistics and mathematical background are provided, to be used in the statistical problem of vocal fold parameters inference (Paper D). Most of the theory here presented is based in the standard textbook (Kaipio and Somersalo, 2005). A summary of the scientific contribution of the thesis are presented at the end of the chapter in relation to a brief summary of the state of the art. Further details can be found in the corresponding paper (Paper D).

4.1 Short introduction to probability

An n -dimensional random variable in a probability space is defined as

$$X : \Omega \rightarrow \mathbb{R}^n, \quad (4.1)$$

denoted by $X \in \mathbb{R}^n$, where Ω is a sample space and $X(\omega) = x \in \mathbb{R}^n$ is a realization of the random variable. Random variables are typically denoted by capital letters, and their realizations by lowercase letters. A probability distribution is generated by this random variable and it is defined as

$$\mu_X(B) = P(X^{-1}(B)) \equiv P(X \in B) \in \mathbb{R} \quad (4.2)$$

with $B \in \mathcal{B}$ the set of events containing the open sets of \mathbb{R}^n and P being a probability measure $P : \mathcal{B} \rightarrow \mathbb{R}$. Random variables that are absolutely continuous define a probability density function $\pi_X : \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$\mu_X(B) = \int_B \pi_X(x) dx, \quad B \in \mathcal{B}. \quad (4.3)$$

The joint probability distribution of two random variables $X_1 \in \mathbb{R}^{n_1}$ and $X_2 \in \mathbb{R}^{n_2}$ is defined as

$$\mu_{X_1 X_2}(B_1, B_2) = P(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)). \quad (4.4)$$

If they are absolutely continuous, then

$$\mu(B_1, B_2) = \int_{B_1 \times B_2} \pi(x_1, x_2) dx_1 dx_2 \quad (4.5)$$

and the marginal probability distribution of X_1 is

$$\mu(B_1) = \int_{B_1 \times \mathbb{R}^{n_2}} \pi(x_1, x_2) dx_1 dx_2, \quad (4.6)$$

which expresses the probability of B_1 regardless of the random variable X_2 . The random variables are independent if

$$\pi_{X_1 X_2}(x_1, x_2) = \pi_{X_1}(x_1) \pi_{X_2}(x_2). \quad (4.7)$$

Given $A, B \in \mathcal{B}$ and $P(B) > 0$, the conditional probability is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (4.8)$$

If random variables are independent, then $\mu(B_1|B_2) = \mu(B_1)$ and the other way around. For the random variables used in this work, it is assumed that the Bayes' formula

$$\pi(x_1|x_2) = \frac{\pi(x_1, x_2)}{\pi(x_2)} = \frac{\pi(x_1|x_2)\pi(x_2)}{\pi(x_2)} = \frac{\pi(x_2|x_1)\pi(x_1)}{\pi(x_2)}, \quad (4.9)$$

is valid.

4.2 Statistical inversion

In Bayesian inversion theory, all variables included in the model are modeled as random variables, and the information related to their realizations is coded in the probability distributions. If $Y \in \mathbb{R}^m$ is the random variable corresponding to the measurement or observable data, and the unknown random variable is $Z \in \mathbb{R}^n$, then the model function f can be expressed stochastically as

$$Y = f(Z, V), \quad (4.10)$$

where the realizations of Y correspond to the data y , and V represents the measurement noise. The solution of a statistical inverse problem is a probability density function conditioned to the knowledge on the observed data. Use of the Bayes' formula yields the posterior probability density function of the unknown random variable Z

$$\pi_{post}(z) = \pi(z|y) = \frac{\pi(z, y)}{\pi(y)} = \frac{\pi_{pr}(z)\pi(y|x)}{\pi(y)} \quad (4.11)$$

where $\pi_{pr}(z)$ is the prior density function which codes the prior knowledge on the unknown variable and $\pi(y|z)$ is the likelihood function. This formula expresses that the prediction is reliable to some degree of probability, after accounting for evidences. In the denominator, the probability density is the marginal density

$$\pi(y) = \int_{\mathbb{R}^n} \pi(z, y) dz = \int_{\mathbb{R}^n} \pi(y|z)\pi_{pr}(z) dz, \quad (4.12)$$

which is a normalization constant, and it is typically disregarded.

Assume now and in the rest of the chapter, that the stochastic model in Eq. (4.10) is linear and that the measurement is corrupted by additive noise, independent of Z . The statistical model can then be expressed as

$$Y = \mathbf{H}Z + V \quad (4.13)$$

for a system matrix \mathbf{H} . Then, the likelihood density is

$$\pi(y|z) = \pi_V(y - \mathbf{H}z). \quad (4.14)$$

Assume further that the prior density function follows a normal distribution, that is, the unknown random variable is $Z \sim \mathcal{N}(z_0, \mathbf{\Gamma}_0)$, and well as the random measurement noise $V \sim \mathcal{N}(v_0, \mathbf{\Gamma}_v)$. The covariance matrices are defined as

$$\mathbf{\Gamma}_0 = \text{cov}[(Z - z_0)(Z - z_0)^T] \quad \text{and} \quad \mathbf{\Gamma}_v = \text{cov}[(V - v_0)(V - v_0)^T], \quad (4.15)$$

which represent the uncertainty of the expected values z_0 and v_0 , and can be estimated for instance given a large sample of realizations. Then, from Eqs. (4.11) and (4.14) it is easy to see that the posterior distribution satisfies

$$\pi_{post}(z) \propto \exp \left(-\frac{1}{2}(z - z_0)^T \mathbf{\Gamma}_0^{-1}(z - z_0) - \frac{1}{2}(y - \mathbf{H}z - v_0)^T \mathbf{\Gamma}_v^{-1}(y - \mathbf{H}z - v_0) \right). \quad (4.16)$$

Because the random variable Y satisfies

$$E[Y] = y_0 = \mathbf{H}z_0 + v_0 \quad (4.17)$$

and

$$\text{cov}[Y] = \mathbb{E}[(Y - y_0)(Y - y_0)^T] = \mathbf{H}\mathbf{\Gamma}_0\mathbf{H}^T + \mathbf{\Gamma}_v \quad (4.18)$$

it can be shown that the expression of the posterior density in Eq. (4.16) can be expressed as (Kaipio and Somersalo, 2005)

$$\pi_{post}(z) \propto \exp\left(-\frac{1}{2}(z - \bar{z})^T \mathbf{\Gamma}_{post}^{-1}(z - \bar{z})\right) \quad (4.19)$$

with

$$\bar{z} = z_0 + \mathbf{\Gamma}_0\mathbf{H}^T(\mathbf{H}\mathbf{\Gamma}_0\mathbf{H}^T + \mathbf{\Gamma}_v)^{-1}(y - \mathbf{H}z_0 - v_0) \quad (4.20)$$

and

$$\mathbf{\Gamma}_{post} = \mathbf{\Gamma}_0 - \mathbf{\Gamma}_0\mathbf{H}^T(\mathbf{H}\mathbf{\Gamma}_0\mathbf{H}^T + \mathbf{\Gamma}_v)^{-1}\mathbf{H}\mathbf{\Gamma}_0 = (\mathbf{\Gamma}_0^{-1} + \mathbf{H}^T\mathbf{\Gamma}_v^{-1}\mathbf{H})^{-1}. \quad (4.21)$$

The mean \bar{z} of the Gaussian posterior distribution, the so-called conditional mean, is a single-value estimate which, given the previous assumptions, coincides with the classical maximum a posteriori estimate

$$z_{MAP} = \arg \max_{z \in \mathbb{R}^n} \pi(z|y) = \bar{z} \quad (4.22)$$

or the maximum likelihood estimate

$$z_{ML} = \arg \max_{z \in \mathbb{R}^n} \pi(y|z) = \bar{z}. \quad (4.23)$$

It is interesting to notice that if the prior knowledge on the unknown variable is normally distributed with zero mean and diagonal covariance matrix $\mathbf{\Gamma}_0 = \gamma^2\mathbf{I}$, and the random noise has also zero mean and covariance $\mathbf{\Gamma}_v = \sigma^2\mathbf{I}$, meaning that there is no correlation between the components of the random variable, then

$$\bar{z} = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T + \lambda\mathbf{I})^{-1}y = (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{I})^{-1}\mathbf{H}^Ty, \quad (4.24)$$

with $\lambda = \sigma^2/\gamma^2$. This expression is equivalent to the classical Tikhonov regularized optical flow solution in Eq. (3.11), which is a maximum-likelihood estimate, as used in Eq. (4.23). Hence, deterministic techniques to overcome ill-conditioning of system matrix when noisy measurements occur, such as Tikhonov regularization, is just a particular case of the general statistical inversion theory, where the advantages of a wise use of the prior knowledge are disregarded. By setting large values of the prior covariance (γ^2 large), that is, decreasing the importance of the prior knowledge, the Tikhonov regularized solution approaches the least-squares solution (λ small), meaning that the solution may be highly contaminated by error in the measurement. Hence, the solution is highly uncertain. Indeed, from Eq. (4.21), large values of the prior covariance lead to a large posterior covariance. It is demonstrated then that a well-justified prior density is a key factor to reduce the uncertainty of the solution. On the contrary, if accurate prior knowledge on the unknown variable is available, the covariance of the prior is highly reduced. Similarly, accurate measurement data decreases the uncertainty of the results, as Eq. 4.21 shows.

4.3 Nonstationary inverse problem

In the previous section the importance of accurate prior knowledge on the unknown variable as well as accurate measurement data has been demonstrated. The prior knowledge may be based on an underlying system corresponding to a prediction model, which in turn carries an error. In many physical applications, such as in biomechanics, the prediction models are time-dependent, and the prior knowledge on the unknown variable depends on previous states of the system. Hence, the statistical inverse problem in Eq. (4.13) becomes nonstationary. Such a situation is modeled by the so-called state-space models (Kaipio and Somersalo, 2005). These models are used in time series analysis (Hamilton, 1994; Frühwirth-Schnatter, 1994), with numerous applications such as macroeconomics and finance forecasting (Chopin, 2004; Creal et al., 2013).

Let $\{Z_k\}_{k=0}^{\infty}$ be an stochastic process formed by the time-dependent unknown variables Z_k , called the state vectors, and $\{Y_k\}_{k=0}^{\infty}$ another stochastic process with the time-dependent observations. It is assumed that the processes have the following properties:

- $\{Z_k\}_{k=0}^{\infty}$ is a p -Markov process: $\pi(z_{k+1}|z_k, z_{k-1}, \dots) = \pi(z_{k+1}|z_{k-p+1}, \dots, z_k)$
- $\{Y_k\}_{k=0}^{\infty}$ is a Markov process with respect to $\{Z_k\}_{k=0}^{\infty}$: $\pi(y_k|z_k, z_{k-1}, \dots) = \pi(y_k|z_k)$
- $\pi(z_{k+1}|z_k, y_k, y_{k-1}, \dots) = \pi(z_{k+1}|z_k)$, so-called Markov transition kernels.

Many physical phenomena can be modeled by a 1-Markov process. Under the assumptions of linearity and Gaussian distributions, a state-space model is

$$Z_{k+1} = \mathbf{G}_{k+1}Z_k + W_{k+1}, \quad \text{state evolution model} \quad (4.25a)$$

$$Y_{k+1} = \mathbf{H}_{k+1}Z_{k+1} + V_{k+1}, \quad \text{observation model} \quad (4.25b)$$

for $k \geq 0$, with additive noises W_{k+1} and V_{k+1} , mutually independent and also mutually independent of the initial state vector Z_0 . The state noise W_{k+1} accounts for the model prediction error. Given a number of observations $y_{1:k} = \{y_1, \dots, y_k\}$, the problem of finding the posterior distribution will depend on the time-dependency between the available measurement and the unknown variable. If the posterior distribution is $\pi(z_k|y_{1:k})$, then the problem is called a filtering problem, and the unknown variables are estimated sequentially. A fixed lag smoothing problem (p -lag smoothing problem) finds the posterior $\pi(z_k|y_{1:k+p})$, where the data can be processes before the estimation is computed. These posterior density generally carry has uncertainty as the probability is conditioned to a larger number of observations (Kaipio and Somersalo, 2005). A prediction problem corresponds to finding the posterior $\pi(z_{k+1}|y_{1:k})$. Each of these problem have their advantages and disadvantages. For instance, the smoothing

problem assures lower uncertainty, but the computational load is high. In this thesis, the algorithm to solve the filtering problem is presented. In addition, a number of parameter estimation algorithm require the solution to a smoothing problem (Paper A). Hence, the smoothing problem is also briefly introduced.

4.3.1 Filtering problem

Given $\pi(z_k|y_{1:k})$, the algorithm to solve the filtering problem of finding the posterior $\pi(z_{k+1}|y_{1:k+1})$ is two-steps:

- Time-evolution updating, which can be seen as prior information: find the prediction $\pi(z_{k+1}|y_{1:k})$ based on the transition kernels $\pi(z_{k+1}|z_k)$ as

$$\pi(z_{k+1}|y_{1:k}) = \int \pi(z_{k+1}, z_k|y_{1:k}) dz_k = \int \pi(z_{k+1}|z_k) \pi(z_k|y_{1:k}) dz_k. \quad (4.26)$$

This probability distribution can be interpreted as $\pi_{pr}(z_{k+1})$, which is useful for the following update. The transition kernels are obtained from the stochastic state evolution model in Eq. (4.25a).

- Observation updating, which can be seen as the posterior: find the filtering $\pi(z_{k+1}|y_{1:k+1})$ given the likelihood $\pi(y_{k+1}|z_{k+1})$, as

$$\pi(z_{k+1}|y_{1:k+1}) = \frac{\pi(y_{k+1}|z_{k+1}) \pi(z_{k+1}|y_{1:k})}{\pi(y_{k+1}|y_{1:k})}, \quad (4.27)$$

with

$$\pi(y_{k+1}|y_{1:k}) = \int \pi(y_{k+1}|z_{k+1}) \pi(z_{k+1}|y_{1:k}) dz_{k+1} \quad (4.28)$$

Similarly to the stationary statistical inverse problem presented in the previous section, the likelihood function $\pi(y_{k+1}|z_{k+1}) = \pi_{V_{k+1}}(y_{k+1} - \mathbf{H}_{k+1} z_{k+1})$ is given by the observation model and follows the same probability distribution as the random measurement error.

As with the stationary statistical inverse problem discussed in Sec. 4.2, if zero mean Gaussian noises $W_{k+1} \sim \mathcal{N}(0, \mathbf{\Gamma}_{w_{k+1}})$ and $V_{k+1} \sim \mathcal{N}(0, \mathbf{\Gamma}_{v_{k+1}})$ are assumed, as well as a Gaussian initial state density distribution $Z_0 \sim \mathcal{N}(z_{0|0}, \mathbf{\Gamma}_{0|0})$, the updating formulas yield Gaussian distributions of the prediction and posterior densities in Eqs. (4.26)-(4.27). Closed expression of the so-called Kalman filters can be found, which provide point estimates of the unknown state vectors corresponding the mean of the posterior density $z_{k+1|k+1} = E[Z_{k+1}|Y_{1:k+1}]$ as well as the covariance $\mathbf{\Gamma}_{k+1|k+1} = \text{cov}[Z_{k+1}|Y_{1:k+1}]$ (Kaipio and Somersalo, 2005). It can be seen that yet the use of well-justified time evolution models with low uncertainty and state or modeling error with small

covariances has a positive effect on reducing the uncertainty of the posterior distribution.

4.3.2 Smoothing problem

The smoothing problem consists of the estimation of the posterior $\pi(z_k|y_{1:k+p})$. If $k+p = N$ is the entire data set, the smoothing problem is called fixed interval smoothing problem. For Gaussian and linear problems, the theory is similar to the filtering problem presented above.

Defining a state-space model for $Z_k = [Z_k, \dots, Z_{k-p}]$, and applying the Kalman filtering formulas, the backward iteration

$$z_{k|N} = z_{k|k} + \mathbf{A}_k(z_{k+1|N} - z_{k+1|k}) \quad (4.29)$$

$$\mathbf{\Gamma}_{k|N} = \mathbf{\Gamma}_{k|k} + \mathbf{A}_k(\mathbf{\Gamma}_{k+1|N} - \mathbf{\Gamma}_{k+1|k})\mathbf{A}_k^T \quad (4.30)$$

for $k = N, \dots, 2$ provides the point estimates, where the backwards gain matrices are

$$\mathbf{A}_k = \mathbf{\Gamma}_{k|k}\mathbf{G}_{k+1}\mathbf{\Gamma}_{k+1|k}^{-1}. \quad (4.31)$$

If \mathbf{G}_k and \mathbf{H}_k are time independent, the computations are drastically simplified.

So far, it has been assumed linearity of the state-space model, and random variables normally distributed. If these assumptions are not fulfilled, point estimates are difficult to be found. Alternatively, Markov chain Monte Carlo methods, known as particle filters in filtering problems, can be used (Chen, 2003; Doucet and Tadić, 2003; Schön et al., 2011). These methods generate a sample distributed according to the posterior density. However, in order to maintain a reasonable computational load, low-dimensional evolution models are required (Hadwin et al., 2016). When a large evolution system matrix must be handled, linearization and the Gaussian assumption drastically simplifies the solution process (Paper D).

4.3.3 State-space identification problem

The statistical inverse problem has been formulated to infer the dynamics of a hidden system, by using knowledge on these dynamics prior to the measurement. Even though a state error W is introduced to assess the uncertainty of the evolution model or prior knowledge, a number of model parameters, typically unknown, are implicit in the model and have a direct effect on the model predictions. In many scientific studies, it is hypothesized that the cause of the observed dynamics of a hidden mechanical system is related to the properties

of the materials, which motivates numerous investigations on parameter estimation (Corigliano and Mariani, 2004; Ching et al., 2006; Schwarz et al., 2008; Hadwin et al., 2016).

The problem of model parameters estimation of a state-space model is known as a state-space identification problem (Kaipio and Somersalo, 2005), Bayesian system identification problem (Beck, 2010), or Bayesian model updating (Ching and Chen, 2007; Simoen et al., 2013). These problems can be solve in multiple ways. For low dimensional system, the parameters are typically modeled as a stochastic process $\{\Theta_k\}_{k=0}^{\infty}$ where Θ_k is the random variable corresponding to a parameter realization θ_k in the matrix \mathbf{G}_k in Eq. (4.25a). Then, a new evolution model can be formulated as

$$Z_{k+1} = \mathbf{G}_{k+1}Z_k + W_{k+1}^1, \quad (4.32a)$$

$$\Theta_{k+1} = \Theta_k + W_{k+1}^2, \quad (4.32b)$$

where the model parameters are assumed to follow a random walk in this case. Then, the filtering and smoothing theory previously described can be applied to the new augmented state variable $(Z_k \ \Theta_k)^T$. Otherwise, particle filter methods can be implemented to approximate the posterior density of the model parameters if non-linearity or non-Gaussian distributions are assumed (Ching et al., 2006; Eftekhari Azam et al., 2012; Hadwin et al., 2016). Otherwise, if linearity and Gaussian distributions are assumed, point estimates can be found by Kalman filtering. However, it is rare that the system depends linearly on the model parameters. Thus, alternative methods must be applied in situations where it is required to manipulate large systems of equations and the application of particle filters is unfeasible.

If the model parameters are assumed to be constants θ and not random variables Θ_k , the problem is typically solved by maximum likelihood algorithms, similarly to Eq. (4.23). The maximum likelihood estimate $\hat{\theta}$ is the value of the unknown model parameters that is most likely to produce the measured data, regardless of the realizations of the underlying system. For mathematical convenience, the log-likelihood is used and the point estimate is found by solving

$$\begin{aligned} \hat{\theta} &= \arg \max_{\theta} \mathcal{L}(\theta; y_{1:N}) = \arg \max_{\theta} \log \pi(y_{1:N}; \theta) \\ &= \arg \max_{\theta} \log \int \pi(y_{1:N}, z_{0:N}; \theta) dz_{0:N}. \end{aligned} \quad (4.33)$$

Finding an expression of the marginal probability density and the solution to this problem is rather a complex task. Alternative algorithms are able to find maximum likelihood estimates by solving an easier problem; for example, the expectation-maximization algorithm (Dempster et al., 1977). Lower bound of the maximum likelihood estimate in Eq. (4.33) can be found in a process described in (Paper D).

4.4 Contributions

Previous studies have focused on vocal fold tissue parameters inference, using direct inverse methods with successful results (Döllinger et al., 2002; Yang et al., 2011; Pinheiro et al., 2012). The lack of uncertainty information and time-varying characteristics has motivated to use of statistical inversion methods (Cataldo et al., 2013; Hadwin et al., 2016), which had been previously used in the vocal tract characteristics inference problem (Kopec, 1986).

In (Cataldo et al., 2013) a non-linear stationary statistical inverse problem is formulated with a two-mass vocal fold model coupled to a vocal tract model, corrupted with Gaussian noise. The vocal fold tension, the neutral glottal area and the subglottal pressure are estimated, along with the uncertainty of the estimation. The output radiated pressure is used as observation data, with special focus on the fundamental frequency of a voice signal. The posterior density is estimated by the importance sampling method, which is a sampling-based method. In (Hadwin et al., 2016) a nonstationary statistical inverse framework is developed. The evolution model is based on a symmetric three-mass body-cover model. Non-linearities are considered by introducing the aerodynamic loading as well as the spring-like collision forces. The vector of displacements of the upper, lower, and body masses, as well as the model parameters (damping and spring coefficients) are estimated. The parameters are assumed to be time-dependent and an augmented state vector as in Eq. (4.32) is used to formulate the state-space model. The observation model is based on simulated glottal areas corrupted by Gaussian noise. The particle filter is used to estimate the parameters.

In an effort to account for the actual complexity of the vocal fold mechanical properties, relevant for clinical purposes, in (Paper D) an observation model based on a continuum mechanical model of the vocal folds is presented. The dimension of the observation vector based on laryngeal videoendoscopic data is increased by means of a dense optical flow algorithm, to compensate for the increasing uncertainty of the feature estimates caused by the large dimension of the unknown continuum state variable. The contribution of the thesis in regard to statistical inversion for vocal fold features estimation is purely theoretical, and due to time limitations no computational results are presented. Further assessment of the presented method is then needed. The contribution can be found in (Paper D), which can be summarized as follows.

Paper D

- A theoretical approach to estimate vocal fold continuum tissue parameters within a statistical state-space framework is presented and discussed.
- An evolution model is based on a collision-free asymmetric continuum

model of the vocal folds with Dirichlet boundary conditions. The observation model is based on a dense optical flow algorithm, thereby increasing the number of observations while avoiding the use of invasive methods.

- Linearity of the evolution model is assumed, neglecting collision forces and the position dependency of the aerodynamic forces. State and measurement noises are assumed to be Gaussian.
- Bayesian filtering methods is used to develop the maximum-likelihood based algorithm.
- The expectation-maximization algorithm is used to find single-valued estimates of the biomechanical parameters by identification of the system matrices as well as the unknown initial probability distributions. To this aim, Kalman filtering and smoothing methods are used.

CHAPTER 5

Overall conclusions

The analytical study of the inference problem of estimating vocal fold features from laryngeal high-speed videoendoscopy resulted in three lines of inquiry for accomplishing this task. The work was presented in four scientific papers and concerned with biomechanical modeling, data acquisition and inversion procedures. In this chapter the overall conclusions are summarized and some directions of further research are discussed.

5.1 Summary of main results

Studies on biomechanical modeling of the vocal folds with a focus on the improvement of the accuracy and precision of the model for inversion purposes and for medical studies were carried out (Paper A). The accuracy was evaluated in the spatial and temporal domain in regard to contact force predictions, also related to the predicted vocal fold displacement field. The precision was evaluated in the context of asymmetric oscillations. This was achieved by developing a vocal fold contact model, and investigating different solution methods and temporal discretization schemes. This line of inquiry responded to and built on previous works on vocal fold modeling. In order to provide a framework suitable for clinical investigations on morphological tissue alterations or vocal fold dysfunction, a transversely isotropic viscoelastic finite element model of

the vocal folds was adapted from the work in (Rosa et al., 2003) and multimass models were discarded. For computational considerations in regard to the inversion process, the flow model was simplified, and a compressive one-dimensional Bernoulli flow was shown to be sufficient to set the complex vocal fold system into oscillation and promote vocal fold separation. The asymmetric contact problem in a finite element model of the vocal fold tissue was investigated and validated. Inspired by earlier work by (Rosa et al., 2003), finite element contact discretization elements were used in the contact model. Differently from the node-to-node discretization presented in (Xue et al., 2012) with prescribed contact location, the node-to-segment contact element presented in this thesis were suitable for intricate finite element meshes and frictional contact, which accommodates arbitrary contact location. In order to avoid a possible spurious effect of contact detection zones based on a minimum gap, a novel contact detection mechanism was presented, which is shown to successfully detect collision at glottal exit. The uncertainty introduced by user-defined spring-like contact forces in terms of dynamical and mechanical accurate predictions was analyzed with different optimization procedures to solve the constrained equations of motion. The Lagrangian approach was shown to produce more accurate results over linear spring-like contact models, and successfully predicted the time-evolution of impact pressure. A smoothing effect of local force prediction was achieved with the augmented Lagrangian method, although yet uncertainty from the parameter choice was introduced (Paper B).

A non-invasive image processing algorithm to acquire data from laryngeal high-speed videoendoscopy was investigated (Paper C). Motivated by the need to reduce the underdeterminess of the continuum vocal fold feature inference problem from *in vivo* laryngeal inspections, a dense optical flow algorithm was adapted to the complex nature of the laryngeal motion, and the two-dimensional full displacement field was extracted. The spurious effect of noise in the images caused by the homogeneity of the intensity field in laryngeal imaging (ill-conditioning) was improved by means of regularization techniques. Discontinuities in the wave-like vocal fold motion were treated with robust statistics, and the effect of the image-sequence temporal discretization was reduced by means of anti-aliasing techniques based on Gaussian spatial smoothing. The results showed that the algorithm is able to successfully capture the vocal fold motion, which was demonstrated by principal components analysis of the estimated optical flow sequence. However, the algorithm fails in evaluating the laryngeal motion at collision. Furthermore, principal components decomposition has been shown to provide information of the hidden vibration pattern. This may help to quantitatively describe laryngeal high-speed videoendoscopy in relation to clinical or scientific studies.

Preliminary theoretical investigations and considerations on the statistical in-

verse problem formulation of vocal fold features inference were presented in Paper D. The problem was formulated in a nonstationary framework, and a state-space model was developed. An evolution model based on a constrained finite element model of the vocal folds was formulated with the aim to investigate physical tissue alterations for clinical research. However, this presented a limitation with regard to the use of particle filter methods for posterior probability sampling. By assuming linearity and a state and measurement errors normally distributed, Kalman filtering methods were applied. An expectation-maximization algorithm was used to solve the state-space identification problem, providing estimates of the system matrices, initial random state and noise distributions. An observation model was developed based on the dense optical flow image sequence. The use of video frame rates not sufficiently large to assure sampling match between time-dependent observation and state vectors motivated the use of an expectation-maximization for missing data. Preliminary results suggested the adequacy of the proposed state-space model, although no computational results were presented, due to time limitations.

5.2 Directions of further research

The work performed in this thesis provides a foundation for future work. This section presents some of these directions of further research.

With regard to the biomechanical model of the vocal folds:

- **Study of the influence of the vocal fold configuration on impact pressure.** Because mass lesions are typically apparent in the anterior half of the vocal folds (Johns, 2003), the anterior-posterior vocal fold asymmetries are hypothesized to play a role on prediction of damage related to vocal fold lesions and phonatory trauma. The effect of such asymmetries could contribute to providing an answer to research questions on correlation between anterior-posterior asymmetries and peak contact stress distribution or frictional contact.
- **Frictional contact.** A necessary future study involves introducing the frictional component of contact forces. This is motivated by the evident anterior-posterior vocal fold motion obtained from the optical flow image sequence (Paper C), and the actual geometrical complexity of the cover layer in both healthy and unhealthy vocal folds, which makes normal contact a weak assumption. Even under a small displacements regime, attention must be paid to the variational formulation.
- **Alternative variational method for energy minimization.** The Lagrange multiplier solution could be regularized by means of the Nitsche

method (Nitsche, 1970) which enforces geometrical constraints in an exact form, and the contact stress is computed from the stress field of the bodies.

- **Non-linear effects on vocal fold collision.** The effect of the subglottal and supraglottal airflow on the vocal fold deformation (Zañartu et al., 2013) when collision occurs could be investigated by nonlinear source-filter methods (Titze and Hunter, 2007).

With regard to optical flow estimation:

- **Further investigations on principal components decomposition of the optical flow image sequence.** Principal components analysis has been shown to successfully distinguish normal and abnormal phonation, by identification of the underlying vibration patterns. Further development of the technique may lead to a fruitful complementary analysis tool for clinical applications, such as classification of functional voice problems that currently lacks visual evidences (Åhlander et al., 2011).
- **Parameter-choice methods for regularization parameter estimation.** With regard to a deterministic optical flow algorithm, parameter-choice techniques, such as L-curve, discrepancy principle or generalized-cross-validation (Hansen, 1998) can be implemented to estimate the amount of noise present in the image sequence and thereby improving the noise effect.
- **Optical flow estimation within a probabilistic framework.** The continuity optical flow equation can be formulated as a stationary inverse problem (Simoncelli et al., 1991; Weiss and Fleet, 2002). This formulation may provide further insights into the uncertainties in the observation model of the state-space model for vocal fold parameter inference.

With regard to statistical inversion methods:

- **Evaluate computational cost and quality of inversion matrices.** The expectation-maximization algorithm based on Kalman filtering and smoothing in a continuum model of the vocal folds requires large matrix multiplication and inversion. Further investigations must be carried out in regard to computational optimization and proper numerical treatment of the matrices.
- **Introduce uncertainties from the camera position.** One of the drawbacks of *in vivo* laryngeal high-speed videoendoscopy is the unknown position of the unstable camera. This increases the uncertainty of velocity magnitude estimates, which may be coded in a camera transformation matrix in the observation model.

- **Investigations on three-dimensional finite-element mesh generation from two-dimensional superior view of the vocal folds.** Another drawback of *in vivo* laryngeal videoendoscopy is that the region of vocal fold tissue captured by the camera is very sensitive to the camera position, as significant part of the vibration can be occluded by the vocal fold cartilages. A very interesting theory is the statistical shape analysis (Cootes et al., 1995), that could be applied to the occlusion problem to infer the regions to further generate the three-dimensional mesh. However, an atlas of the laryngeal geometry is required.
- **Simulations and validation of the state-space identification model is necessary.** Synthetic vocal fold models (Murray and Thomson, 2012; Lodermeier et al., 2015) can be used to a first approach to test the statistical algorithm for vocal fold parameters inference.
- **Application to patient data.** Given the validity of the algorithm, laryngeal high-speed videoendoscopic data from patients with functional voice problems could be used to test tissue-related hypotheses on the injury processes that current quantitative analysis tools failed at evaluating (Åhlander et al., 2011). Objective clinical methods could then be provided, so that one can distinguish individuals with normal voice health from those with functional voice problems.

Journal and conference papers

Paper A

A numerical strategy for finite element modeling of frictionless asymmetric vocal fold collision

Authors: Alba Granados, Marek Krzysztof Misztal, Jonas Brunskog, Vincent Visseq, and Kenny Erleben

Published online in: *International Journal for Numerical Methods in Biomedical Engineering*, 2016. doi: 10.1002/cnm.2793cnm

A numerical strategy for finite element modeling of frictionless asymmetric vocal fold collision

Alba Granados^{1*}, Marek Krzysztof Misztal², Jonas Brunskog¹, Vincent Visseque³
and Kenny Erleben⁴

¹*Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby DK-2800, Denmark*

²*Niels Bohr Institute, University of Copenhagen, Copenhagen DK-2100, Denmark*

³*Institut Supérieur de Mécanique de Paris, Saint-Ouen F-93400, France*

⁴*Department of Computer Science, University of Copenhagen, Copenhagen DK-2100, Denmark*

SUMMARY

Analysis of voice pathologies may require vocal fold models that include relevant features such as vocal fold asymmetric collision. The present study numerically addresses the problem of frictionless asymmetric collision in a self-sustained three-dimensional continuum model of the vocal folds. Theoretical background and numerical analysis of the finite-element position-based contact model are presented, along with validation. A novel contact detection mechanism capable to detect collision in asymmetric oscillations is developed. The effect of inexact contact constraint enforcement on vocal fold dynamics is examined by different variational methods for inequality constrained minimization problems, namely the Lagrange multiplier method and the penalty method. In contrast to the penalty solution, which is related to classical spring-like contact forces, numerical examples show that the parameter-independent Lagrange multiplier solution is more robust and accurate in the estimation of dynamical and mechanical features at vocal fold contact. Furthermore, special attention is paid to the temporal integration schemes in relation to the contact problem, the results suggesting an advantage of highly diffusive schemes. Finally, vocal fold contact enforcement is shown to affect asymmetric oscillations. The present model may be adapted to existing vocal fold models, which may contribute to a better understanding of the effect of the non-linear contact phenomenon on phonation.

Copyright © 2016 John Wiley & Sons, Ltd.

Received 9 October 2015; Revised 23 February 2016; Accepted 28 March 2016

KEY WORDS: biomechanics, vocal folds, contact detection, impact problem, variational methods, computer model.

1. INTRODUCTION

A large number of vocal fold models have been studied in the last sixty years; from simplistic lumped-mass models driven by an incompressible Bernoulli flow [1], to continuum models coupled to advanced Navier-Stokes turbulent flows [2]. The development of vocal fold models is typically motivated by voice production or clinical research. For purposes of clinical research, studies on the mechanical and dynamical conditions that arise during vocal fold contact are of special interest, specially tissue stresses and reaction forces [3]. In that context, a three-dimensional finite element model of the vocal folds provides the framework for a detailed deformation analysis [4, 5]. In order to perform a full investigation of the vocal fold deformation,

*Correspondence to: Technical University of Denmark, Ørstedes Plads, Building 352, Kgs. Lyngby DK-2800, Denmark. E-mail: albagranados@gmail.com

the model may include flow-structure interaction to induce vocal fold separation after contact. Furthermore, vocal fold oscillations are not generally symmetric. Therefore, allowance of asymmetric contact is of great importance.

Previous research on self-sustained vocal fold models typically include collision. In an early model presented in [6] a simple one-dimensional lumped-mass model of the vocal folds driven by an incompressible Bernoulli flow simulates collision by spring-like reaction forces. However, the simplicity of the model does result in it lacking desirable features for pathological vocal fold investigations, such as highly detailed internal stress estimations and high spatial resolution. Similar limitations may be argued in relation to the self-sustained lumped-mass model with a symmetric Hertzian contact model in [7]. In order to overcome these limitations, a finite element continuum model of the tissue is developed in [4]. The model is used to analyze the mechanics and dynamics at vocal fold closure. Symmetric vocal fold contact is assumed and the impact forces are orthogonal to a virtual rigid middle-sagittal plane that prevents node interpenetration. Because airflow coupling is not included in the model, the contact phase of vocal fold separation is not investigated. Furthermore, the study limits the analysis to the mechanical output of the model, and it does not provide numerical details of the contact model. Analogous model features are found in [8], where the symmetric Hertzian contact in [7] is adapted to a continuum model of the vocal folds. Advanced models with full fluid-structure coupling typically assume symmetric vocal fold vibrations. In these cases, contact forces are modeled as spring-like forces [9], or computed by means of an Augmented Lagrangian algorithm [10], with no specific reference to contact model parameters. Furthermore, fully coupled numerical fluid-structure solvers may present topological limitations in relation to the flow solver. Hence, a minimum glottal gap is typically enforced to model the contact interaction between the two vocal folds [11], or the collision phenomenon is simply disregarded [2].

The asymmetric contact problem in self-sustained vibrations of the vocal folds is typically disregarded in continuum models of the tissue [12], or formulated in simple lumped-mass models [6, 13]. As pointed out in [9], sophisticated contact detection algorithms are required for asymmetric oscillations and intricate vocal folds geometries, specially at the upper vocal fold edges where convex contacting surfaces are present [14]. However, a recent continuum fluid-structure model is capable of detecting asymmetric contact [15]. The contact interface constitutive equations are described on a nodal basis following a non-linear, spring-based law. A minimum glottal gap defines the contact detection zone. However, the node-to-node formulation reported in the study is not valid for general mesh generators that typically create non-matching meshes, because it does not allow for tangential and normal decomposition [16]. Furthermore, specific model contact parameters are necessary to obtain realistic vocal fold collision, but adequate experimental data is currently unavailable; this may explain the lack of validation of the vocal fold contact model in [13]. Alternatively, a more general contact model is developed in [17]. In their work, a classical combinatorial search between mesh elements at the two vocal folds is performed, despite the computational load of this operations. The contact model could potentially allow for asymmetric contact and provide detailed simulations of the impact forces. However, these features are not investigated, and the contact problem lacks validation. Furthermore, the results presented in their study seem to be inaccurate in terms of displacement, as surface contact nodes penetrate the tissue. This behavior may be a consequence of the numerical method used to enforce contact; however, the description of the numerical algorithm is missing in the study. Some vocal fold models enforce the contact constraint by the penalty method, which is equivalent to adding linear spring forces at contact [6, 9], or an augmented Lagrangian algorithm [10]. However, all these solution algorithms require a thoughtful choice of the model parameters, which makes the model output sensitive to arbitrariness. A recent investigation has shown that the dynamical behavior of a lumped-mass vocal fold model is highly sensitive to simplifications in the contact formulation [18]. These simplifications might lead to biased interpretation of the mechanical properties at contact, specially relevant in voice pathology diagnose. Finally, accuracy in contact dynamics estimation is of importance in inverse problems for vocal fold tissue parameters estimation [8].

In this paper, a three-dimensional viscoelastic continuum vocal fold model that includes a finite element model of vocal fold collision is developed. Vocal fold closure and separation is induced by an incompressible Bernoulli flow. The collision is modeled by a position-based inequality constraint, implying a constitutive contact law and a specific mechanical behavior. Different numerical procedures are adopted from contact mechanics to the vocal fold contact problem. A contact discretization technique that allows for vocal fold sliding is presented, suitable for non-matching body meshes. Vocal fold contact is detected by a novel mechanism based on deformable interface tracking. The algorithm is capable to detect contact in asymmetric oscillations and at convex surfaces such as the upper glottal boundary, and it avoids classical combinatorial contact search. Furthermore, it is hypothesized that non-physical vocal fold interpenetration or minimum glottal gap enforcement affects the predicted dynamical and mechanical behavior of the system. Hence, the effect on vocal fold dynamics of simplifications in the contact model with regard to position-based constraints (e.g., see [13, 11]) are investigated via the penalty solution algorithm. The results are compared with the proposed Lagrange multiplier formulation for exact contact constraint enforcement. Furthermore, given the high kinetic energy that the vocal folds contain just previous to collision and the short vocal fold impact time, linear spring-like contact forces are conjectured to fail at distributing the energy of impact. Hence, the penalty and the Lagrange multiplier solutions here investigated allow for a deeper understanding of the contact mechanism. Finally, the present paper investigates the effect of temporal discretization schemes on the numerical solution to vocal fold mechanics at collision.

The outline of the paper is as follows. In Section 2 the methodology to model the airflow, tissue and collision between vocal folds is presented, including numerical aspects. Section 3 is devoted to the setup for the simulation results, followed by Section 4, which includes model validation, results and discussion of the present contact model. Finally, in Section 5 the conclusions are presented.

2. METHODOLOGY

A three-dimensional self-sustained finite element model of the vocal folds with contact is developed. The airflow model is based on the one-dimensional models in [1, 19], and it is adapted to the three-dimensional geometry of the intraglottal region. The material model is based on previous works [5, 17]. The collision problem between the vocal folds is investigated, including mathematical formulations, the contact discretization and the collision detection mechanism. Furthermore, an analysis of the temporal discretization techniques is included to investigate the model performance at non-smooth contact.

2.1. Glottal airflow model

For an incompressible fluid with constant density ρ_{air} , the Bernoulli flow at a given time through a spatially discretized glottal canal along the longitudinal z -axis is described as [19]

$$p(z_{i+1}) = p(z_i) + \frac{\rho_{air}}{2} U^2 \cdot \left(\frac{1}{a(z_i)^2} - \frac{1}{a(z_{i+1})^2} \right), \quad (1)$$

where U is the volumetric flow rate, $p(z_i)$ and $p(z_{i+1})$ are aerodynamic pressures, and average flow speeds over the transverse cross-sectional areas $a(z_i)$ and $a(z_{i+1})$ have been assumed; see Figure 1. Note that the aerodynamic pressures implicitly depend on time since cross-sectional areas depend on the deforming tissue boundary. No energy loss is considered neither at glottal inlet nor at glottal outlet [20]. Assuming a negligible subglottal kinetic energy, $a(z_{sub}) \gg a(z_i)$, as well as no vocal tract coupling, that is, the supraglottal aerodynamic pressure is zero, from Eq. (1) the volumetric flow rate may be approximated as

$$U = \sqrt{\frac{2p_{sub}}{\rho_{air}}} a(z_{det}), \quad (2)$$

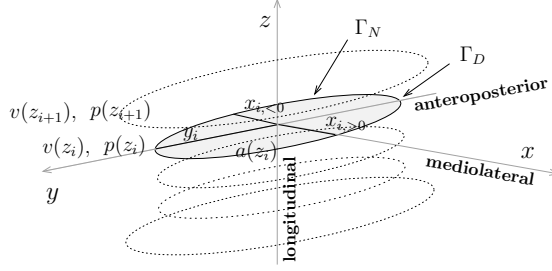


Figure 1. Schematic representation of the larynx assuming elliptical shaped cross-sections. Subscripts indicate spatial discretization along the longitudinal axis. $p(z_i)$, $v(z_i)$ and $a(z_i)$ are the aerodynamic pressure, average flow velocity and cross-sectional area. y_i is the major axis in the anteroposterior direction, and $x_{i,>0}$ and $x_{i,<0}$ are the positive and negative minor axes in the mediolateral direction, respectively. Γ_N and Γ_D indicate the Neumann and Dirichlet boundary, respectively.

where p_{sub} is the subglottal pressure, and $a(z_{det})$ is the cross-section where the flow detaches from the vocal folds boundaries. Previous studies have investigated flow separation in vocal folds [21]. The flow separation point is an important feature for realistic voice simulations. Since the present model aims at resolving vocal fold mechanics, for the sake of simplicity the flow is assumed to detach from the boundary at the minimum cross-sectional area $a(z_{det}) = a_{min}$; downstream that point, a jet is formed, and aerodynamic forces are zero.

By spatial discretization of the vocal fold boundary, the transversal areas $a(z_i)$ in Eq. (1) are computed by assuming an elliptical shape, as in [17]; see Figure 1. In the present study, the computations take into account asymmetrical mediolateral deformation of the vocal folds by allowing for independent minor axis, that is,

$$a(z_i) = \frac{\pi}{2} y_i (x_{i,>0} + x_{i,<0}), \quad (3)$$

where y_i is the major axis in the anteroposterior direction, and $x_{i,>0}$ and $x_{i,<0}$ are the positive and negative minor axes in the mediolateral direction.

2.2. Continuum mechanical model for the vocal folds

2.2.1. Kinematics Under deformation of a body a material particle is displaced from its original position $\mathbf{X} \in V_{solid} \subset \mathbb{R}^3$ (material coordinates) to its current position $\varphi(\mathbf{X}) = \mathbf{x} \in v_{solid} \subset \mathbb{R}^3$ (spatial coordinates) via the deformation mapping $\varphi: V_{solid} \rightarrow \mathbb{R}^3$. The displacement field $\mathbf{u}(\mathbf{X}, t) = \mathbf{x} - \mathbf{X}$ is thereby determined. From basic continuum mechanics, the motion of the body in the absence of volume forces is described by

$$\nabla \cdot \boldsymbol{\sigma} = \rho \frac{\partial^2 \mathbf{x}}{\partial t^2} \quad \text{in } v_{solid}, \quad (4)$$

where ρ is the spatial mass density, and $\boldsymbol{\sigma}$ is the Cauchy stress tensor. In the case of the vocal folds, the Cauchy stress tensor at the moving airflow boundary $\Gamma_N \subset \partial v_{solid}$ with outward normal \mathbf{n} satisfies the Neumann boundary condition

$$\boldsymbol{\sigma} \cdot \mathbf{n} = \mathbf{p} \quad \text{on } \Gamma_N, \quad (5)$$

where $\mathbf{p} = p(z_i)\mathbf{n}$ from Eq. (1). Furthermore, Dirichlet boundary conditions enforce zero displacement at the anteroposterior glottal extreme regions $\Gamma_D \subset \partial v_{solid}$.

It is widely accepted in the voice research community, that vocal folds undertake small displacements during normal phonation [5]. In the present model, the small deformations assumption is also made, and linear theory of deformation is applied. The strain tensor

is approximated by the Cauchy's strain tensor ϵ . The laryngeal tissue is modeled with a compressible viscoelastic Kelvin-Voigt model as in [17], with dynamic viscosity η and stiffness matrix [22]

$$\mathcal{S} = \begin{pmatrix} \frac{(E' - \nu'^2 E)E}{(1+\nu)k} & \frac{EE'\nu'}{k} & \frac{(\nu'^2 E + \nu E')E}{(1+\nu)k} & 0 & 0 & 0 \\ \frac{EE'\nu'}{k} & \frac{E'^2(1-\nu)}{k} & \frac{EE'\nu'}{k} & 0 & 0 & 0 \\ \frac{(\nu'^2 E + \nu E')E}{(1+\nu)k} & \frac{EE'\nu'}{k} & \frac{(E' - \nu'^2 E)E}{(1+\nu)k} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu' & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu' & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix}, \quad (6)$$

where E and ν are the transversal Young's modulus and Poisson's ratio, respectively, E' , μ' and ν' correspond to the longitudinal Young's and shear modulus and Poisson's ratio, respectively, and $k = E'(1 - \nu) - 2\nu'^2 E$. The constitutive equation is

$$\sigma(t) = \mathcal{S}\epsilon(t) + \eta \frac{\partial}{\partial t} \epsilon(t), \quad (7)$$

where symmetry of the stress and strain tensors is assumed.

For admissible test functions \mathbf{w} that vanish on Γ_D with square-integrable derivatives [23], the problem takes the suitable weak form

$$\int_{v_{\text{solid}}} \rho \mathbf{w}^T \cdot \frac{\partial^2 \mathbf{x}}{\partial t^2} dv + \int_{v_{\text{solid}}} \nabla \mathbf{w}^T : \sigma dv = \int_{\Gamma_N} \mathbf{w}^T \cdot \mathbf{p} d\Gamma \quad (8)$$

used as a starting point for the finite element formulation. In analytic mechanics, the equilibrium equations or the weak form in Eq. (8) emanate from the Hamilton's principle, that is, through the process of minimizing a functional $\Pi(\mathbf{v})$ that depends on a displacement field \mathbf{v} and involves the potential and kinetic energies [24, 16],

$$\Pi(\mathbf{v}) = \int_{v_{\text{solid}}} \rho \mathbf{v}^T \frac{\partial^2 \mathbf{v}}{\partial t^2} dv + \int_{v_{\text{solid}}} \nabla \mathbf{v}^T : \sigma dv - \int_{\Gamma_N} \mathbf{v}^T \cdot \mathbf{p} d\Gamma. \quad (9)$$

In this sense, Eq. (8) is called the variational equation, and $\mathbf{w} = \delta \mathbf{v}$ are the variations of the displacements or virtual displacements in the engineering literature. In this study, the vocal fold motion is found as the solution to a minimization problem with the functional in Eq. (9) being the objective function.

2.2.2. Contact constraint When vocal fold collision occurs, additional boundary conditions on the contact surface $\Gamma_C \subset \partial v_{\text{solid}}$ arise to avoid body interpenetration, leading to a variational term C_C which must be added to the left-hand side of Eq. (8). As the large displacements take place in the mediolateral direction [17], large relative tangential movements are not expected to occur in the contact interface. Therefore, frictionless contact is assumed in this work, even though the following formulation can be easily extended to frictional contact. The non-penetration constraint may be expressed in the current configuration v_{solid} in terms of the normal gap g_N in the classical master-slave configuration as [16]

$$g_N = (\mathbf{x}^s - \bar{\mathbf{x}}^m) \cdot \bar{\mathbf{n}}^m \geq 0. \quad (10)$$

The superscript s corresponds to the interface slave node that penetrates into the master surface denoted by the superscript m . The vector $\bar{\mathbf{n}}^m$ is the outward normal on the master surface at the point $\bar{\mathbf{x}}^m$, which is the projection of \mathbf{x}^s onto the master surface; see Figure 2 for an illustration. When contact occurs, that is, when $g_N = 0$, a Cauchy traction vector $\bar{\mathbf{t}}$ is active at the contact point. This reaction force must satisfy the action-reaction principle, and for non-adhesion contact its normal component $r_N = \bar{\mathbf{t}} \cdot \bar{\mathbf{n}}^m$ must be compressive. These

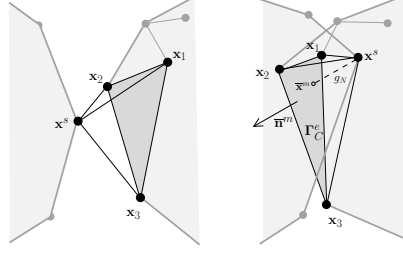


Figure 2. Schematic representation of the conforming interface mesh for collision detection; transverse view before collision (left), after collision (right). Γ_C^e indicate the surface boundary of the node-to-surface contact element.

conditions may be written in a compact form as

$$g_N \geq 0, \quad r_N \leq 0, \quad g_N r_N = 0 \quad \text{on } \Gamma_C, \quad (11)$$

which are referred to as primal feasibility, dual feasibility and complementary slackness, respectively, in Lagrangian duality theory. Nevertheless, the classical master-slave configuration that has been described requires two implementation issues to be taken into consideration in the vocal fold problem. Firstly, collision at the glottal exit may not be detected due to the convexity of the tissue surface boundary, where non-existent or positive normal gaps may be misleading. Secondly, the geometrical and mechanical symmetry of the system makes the master and slave choice arbitrary. As arbitrariness in the master-slave configuration may have a negative effect on the solution, all slaves nodes must be tested against all surfaces, increasing thereby the computational time. In later sections, a novel detection mechanism is presented, which arises as a solution to the problems that classical master-slave implementation may cause.

The branch of mathematics of calculus of variations provides multiple methodologies to solve a constrained nonlinear minimization problem, such as the present. In this work, the Lagrange multiplier and the penalty methods based on [16] are investigated. These methods can be applied to enforce the collision constraints in Eq. (10) in an exact or a relaxed way, respectively. The Lagrange multiplier solution is here proposed as an alternative to the inexact vocal fold collision solutions typically found in the existing literature. In the next paragraphs, the new variational equations are described.

The Lagrangian approach consists of enforcing non-penetration by appending a functional

$$\Pi(\mathbf{v}, \lambda)^{LM} = \int_{\Gamma_C} \lambda g_N(\mathbf{v}) \, d\Gamma \quad (12)$$

to the energy functional $\Pi(\mathbf{v})$ in Eq. (9), with λ being a scalar parameter called the Lagrange multiplier or dual variable, and g_N is the normal gap in Eq. (10). The resulting functional is called the Lagrangian function. The Lagrange multiplier can be seen as the non-positive reaction pressure r_N necessary to avoid body interpenetration, that is, $r_N = \lambda < 0$. Variation of the functional leads to the variational term

$$C_C^{LM} = \int_{\Gamma_C} \lambda (\mathbf{w}^s - \bar{\mathbf{w}}^m) \cdot \bar{\mathbf{n}}^m \, d\Gamma + \int_{\Gamma_C} \delta \lambda (\mathbf{x}^s - \bar{\mathbf{x}}^m) \cdot \bar{\mathbf{n}}^m \, d\Gamma \quad (13)$$

where $\delta \lambda$ refers to the virtual Lagrange multiplier. The above term must be combined with Eq. (8), and together with conditions in Eq. (11), they constitute the Kuhn-Tucker-Karush necessary conditions for optimality [25]. One of the drawbacks of the Lagrange method is the

increase of the dimension of the system by addition of unknowns in the form of Lagrange multipliers. Furthermore, necessary conditions for the solution to be optimal lead to an indefinite system matrix, that is, critical points occur at saddle points [25].

In contrast to the exact contact boundary condition enforcement method previously described, allowance of the vocal fold interpenetration at collision is widely used in vocal fold models [6, 17, 9]. This interpenetration is a consequence of a penalty method approach to the constrained minimization problem, which corresponds to spring-like reaction force. The method consists of the approximation of the Lagrange multiplier, or reaction pressure, as $\lambda \approx \kappa g_N$, with κ a positive scalar that controls how severely the constraint is violated. The energy functional related to the contact interface is

$$\Pi(\mathbf{v})^P = \int_{\Gamma_C} \frac{1}{2} \kappa (g_N(\mathbf{v}))^2 d\Gamma. \quad (14)$$

Variation of this functional yields

$$C_C^P = \int_{\Gamma_C} \kappa (\mathbf{w}^s - \bar{\mathbf{w}}^m) \cdot (\mathbf{x}^s - \bar{\mathbf{x}}^m) d\Gamma, \quad (15)$$

which must be combined with Eq. (8). The penalty method maintains the dimension of the collision-free problem. Enforcing the non-penetration constraint by increasing the penalty parameter κ derives in an ill-conditioning of the global matrix, and collision forces with no physical sense. Therefore, the user is required to make a thoughtful choice of the penalty parameter. Numerical solutions of the Lagrange multiplier and penalty approaches for vocal fold contact are investigated in Section 4.2.

2.3. Discretizations

2.3.1. Spatial discretization The spatial discretization of the solid is performed in a simple way typically found in the literature (e.g., see [5]). Considering the spatial finite element discretization of the solid domain in n_{el} elements and n_{np} nodal points, the equilibrium Eq. (8) yields the matrix form

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}(\mathbf{x} - \mathbf{X}) = \mathbf{F}, \quad (16)$$

where \mathbf{M} , \mathbf{C} , and \mathbf{K} are the mass, damping, and stiffness $3n_{np} \times 3n_{np}$ matrices, respectively, \mathbf{F} is the $3n_{np}$ -vector of applied aerodynamic forces, and \mathbf{x} and \mathbf{X} are the $3n_{np}$ -vectors that contain the current and the original positions of the nodal points, respectively; dots indicate time derivatives. By decomposition of the test functions \mathbf{w} in a finite number of linear interpolation functions defined in a tetrahedral mesh, the system matrices in Eq. (16) can be easily derived by means of isoparametric transformations [23].

The aerodynamic forces in Eq. (16) are derived from the extension of the airflow spatial discretization of the one-dimensional problem in [6] to the present three-dimensional solid boundary. The linear aerodynamic forces below the flow separation point acting on the three vertices of a triangular element face $\partial v_{\text{solid}}^e \subset \Gamma_N$ with current positions $\mathbf{x}_i = (x_i, y_i, z_i)$, $\mathbf{x}_j = (x_j, y_j, z_j)$, and $\mathbf{x}_k = (x_k, y_k, z_k)$ satisfy

$$\mathbf{F}_l = -\mathbf{n} \cdot \begin{cases} \frac{\Delta^e}{6} \left(p(z_i) + \frac{p(z_j)}{2} + \frac{p(z_k)}{2} \right) & \text{if } l = i \\ \frac{\Delta^e}{6} \left(\frac{p(z_i)}{2} + p(z_j) + \frac{p(z_k)}{2} \right) & \text{if } l = j \\ \frac{\Delta^e}{6} \left(\frac{p(z_i)}{2} + \frac{p(z_j)}{2} + p(z_k) \right) & \text{if } l = k, \end{cases} \quad (17)$$

where $p(z_l)$ correspond to the pressure derived from Eq. (1) applied to each nodal point \mathbf{x}_i , \mathbf{x}_j and \mathbf{x}_k , and Δ^e stands for the area of the triangular element face with outward normal vector \mathbf{n} . Downstream the flow separation point, aerodynamic forces are zero

When collision occurs, the finite element discretization of Eq. (13) or Eq. (15) must be calculated. A node-to-surface contact element may be described by a master element surface, defined by target vertices \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 , and a contactor slave node \mathbf{x}^s , belonging to the opposite vocal fold; see Figure 2 (right). The triangular interface geometry can be parametrized by an isoparametric transformation with linear functions $N_I(\xi, \zeta)$ and local variables (ξ, ζ) defined on a reference triangular element with vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$. Hence, the displacement constraint in Eq. (10) yields

$$\left(\mathbf{x}^s - \sum_{I=1}^3 N_I(\bar{\xi}, \bar{\zeta}) \mathbf{x}_I \right) \cdot \bar{\mathbf{n}}^m \geq 0, \quad (18)$$

with $(\bar{\xi}, \bar{\zeta})$ satisfying

$$\left(\mathbf{x}^s - \sum_{I=1}^3 N_I(\bar{\xi}, \bar{\zeta}) \mathbf{x}_I \right) \cdot (\mathbf{x}_k - \mathbf{x}_1) = 0 \quad (19)$$

for $k = 2, 3$, which corresponds to the equations to find the projection $\bar{\mathbf{x}}^m$ of \mathbf{x}^s onto the master surface. Consequently, prescribed position constraints can be expressed by a contact element matrix

$$\mathbf{g}^e = \left(\bar{\mathbf{n}}^m - N_1(\bar{\xi}, \bar{\zeta}) \bar{\mathbf{n}}^m - N_2(\bar{\xi}, \bar{\zeta}) \bar{\mathbf{n}}^m - N_3(\bar{\xi}, \bar{\zeta}) \bar{\mathbf{n}}^m \right) \in \mathbb{R}^{1 \times 12}, \quad (20)$$

with $\mathbf{g}^e \cdot (\mathbf{x}^s - \mathbf{x}_1 - \mathbf{x}_2 - \mathbf{x}_3)^T \geq 0$. Note that for linear shape functions the outward normal is independent of the projection point. For n_{ce} contact elements with negative normal gap, a finite element assembly algorithm can be used to build a global constraint contact $n_{ce} \times 3n_{np}$ matrix \mathbf{G} .

Enforcing the contact displacement constraint with the Lagrangian approach in a finite element formulation, requires the spatial discretization of the variational term in Eq. (13). Interpretation of the Lagrange multipliers as reaction forces, makes it possible to remove the surface integrals in Eq. (13). Hence, the equations that provide the current equilibrium position with n_{ce} active contact elements are

$$\begin{aligned} \mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}(\mathbf{x} - \mathbf{X}) + \mathbf{G}^T \boldsymbol{\Lambda} &= \mathbf{F} \\ \mathbf{G}\mathbf{x} &= \mathbf{0}, \end{aligned} \quad (21)$$

where $\boldsymbol{\Lambda} = (\lambda_1, \dots, \lambda_{n_{ce}})^T$ is a n_{ce} -vector that contains non-positive reaction forces, and the $3n_{np}$ -vector $\mathbf{G}^T \boldsymbol{\Lambda}$ contains the applied contact force at each degree of freedom under contact. For positive Lagrange multipliers, the corresponding contact element must be removed from the global constraint contact matrix to allow for contact surface separation.

By spatial finite element discretization of the variation in Eq. (15), the equilibrium position \mathbf{x} that fulfills the contact constraint in a relaxed way via a penalty formulation satisfies

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}(\mathbf{x} - \mathbf{X}) + \mathbf{G}^T \kappa \mathbf{G}\mathbf{x} = \mathbf{F}, \quad (22)$$

where the penalty parameter κ accounts for the area of the master surfaces, so that the surface integral in Eq. (15) is neglected. The penalty term can be interpreted as a reaction force normal to the master surface and proportional to the contact element volume interpenetration. From this observation, computations can be simplified, as will be shown in Section 2.4.

Enforcement of Dirichlet boundary conditions at n_d nodal points on Γ_D may be expressed in terms of a global solid boundary $n_d \times 3n_{np}$ matrix \mathbf{D} as

$$\mathbf{D}(\mathbf{x} - \mathbf{X}) = \mathbf{0}. \quad (23)$$

Since the constraint matrix has to be assembled only once, the Lagrangian approach with a n_d -vector of unknowns $\boldsymbol{\Lambda}^D$ can be used to modify the equilibrium equation in Eq. (16) as in Eq. (21) to enforce Dirichlet boundary conditions. For equality constraints as in Eq. (23), the Lagrange multipliers can take on both positive and negative values.

2.3.2. Temporal discretization To solve for the equilibrium current position \mathbf{x} , attention has to be paid to temporal discretization schemes, since instabilities and accuracy problems may arise. The accuracy analysis found in most of the literature assumes smoothness of the system, as well as small Rayleigh damping [23]. However, vocal fold contact modeling results in a non-smooth system. Simple lumped-mass models of the vocal folds include a smooth contact activation function [6], which makes classical accuracy analysis of the temporal integration schemes valid. The present finite element vocal fold model is highly non-linear and non-smooth, and due to its complexity the use of smooth activation functions requires alternative contact detection mechanism. Hence, the effect of the temporal schemes in the present system requires computer evaluation. In this section, temporal integration algorithms and its basic features are summarized and adapted to the vocal fold contact equations.

The most common schemes to temporally discretize the hyperbolic problem belong to the Newmark family of implicit one-step methods [26], as used in [17]. These schemes consist of the equations

$$\mathbf{x}_{n+1} = \underbrace{\mathbf{x}_n + h\dot{\mathbf{x}}_n + \frac{h^2}{2}(1-2\beta)\ddot{\mathbf{x}}_n + \beta h^2\ddot{\mathbf{x}}_{n+1}}_{\tilde{\mathbf{x}}_{n+1}}, \quad (24a)$$

$$\dot{\mathbf{x}}_{n+1} = \underbrace{\dot{\mathbf{x}}_n + (1-\gamma)h\ddot{\mathbf{x}}_n}_{\tilde{\dot{\mathbf{x}}}_{n+1}} + \gamma h\ddot{\mathbf{x}}_{n+1}, \quad (24b)$$

where now the subscript n indicates that the function is approximated at time t_n , h is the time step, γ and β are the algorithm parameters associated to the quadrature scheme, and $\tilde{\mathbf{x}}_{n+1}$ and $\tilde{\dot{\mathbf{x}}}_{n+1}$ are the so-called predictors. The algorithm parameters control accuracy and stability of the scheme, which can be studied by spectral analysis of the amplification matrix [23]. For smooth systems, and if Rayleigh damping is assumed, the conditions $2\beta \geq \gamma \geq \frac{1}{2}$ (if equality, it corresponds to an undamped Newmark scheme, as used in e.g., [9]) and $\beta = (\gamma + 1/2)^2/4$ with $\gamma > 1/2$ assures unconditional stability. The latter values introduce numerical dissipation in all frequency range, specially at higher modes, where typically finite element solution seems to deteriorate. Therefore, the scheme may be considered advantageous. However, the accuracy of the algorithm drops to first order.

An alternative method with second-order accuracy and numerical damping at high frequencies is the Hilber-Hughes-Taylor α -method [27]. The scheme consists of a weighted mean of the Newmark approximations in Eqs. (24) as $(1 + \alpha)\mathbf{x}_{n+1} - \alpha\mathbf{x}_n$ and $(1 + \alpha)\dot{\mathbf{x}}_{n+1} - \alpha\dot{\mathbf{x}}_n$, respectively. When α is zero, the scheme is equivalent to the Newmark family of methods. The weight parameter α controls the amount of numerical dissipation; the smaller, the more mode damping. It has been shown that for $\alpha \in [-1/3, 0]$, $\gamma = 0.5 - \alpha$ and $\beta = (1 - \alpha)^2/4$, the scheme is unconditionally stable with second-order accuracy. Nevertheless, the scheme introduces a phase error larger than the one produced by an undamped Newmark rule, which increases as the parameter α is decreased.

In the case of the Lagrange multiplier formulation of the constrained problem, the temporal integration scheme leads to the system of linear equations

$$\begin{bmatrix} \mathbf{A} & \mathbf{D}^T & \mathbf{G}_{n+1}^T \\ (1+\alpha)h^2\beta\mathbf{D} & \mathbf{0} & \mathbf{0} \\ (1+\alpha)h^2\beta\mathbf{G}_{n+1} & \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \ddot{\mathbf{x}}_{n+1} \\ \Lambda_{n+1}^D \\ \Lambda_{n+1} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_n - \mathbf{C} \left((1+\alpha)\tilde{\mathbf{x}}_{n+1} - \alpha\dot{\mathbf{x}}_n \right) - \mathbf{K} \left((1+\alpha)\tilde{\mathbf{x}}_{n+1} - \alpha\mathbf{x}_n - \mathbf{X} \right) \\ -\mathbf{D} \left((1+\alpha)\tilde{\mathbf{x}}_{n+1} - \alpha\mathbf{x}_n - \mathbf{X} \right) \\ -\mathbf{G}_{n+1} \left((1+\alpha)\tilde{\mathbf{x}}_{n+1} - \alpha\mathbf{x}_n \right) \end{bmatrix}, \quad (25)$$

with

$$\mathbf{A} = \mathbf{M} + (1 + \alpha)\gamma h\mathbf{C} + (1 + \alpha)h^2\beta\mathbf{K}. \quad (26)$$

The global contact constraint matrix \mathbf{G}_{n+1} is built from active contact elements at time t_{n+1} , and \mathbf{F}_n is the vector of aerodynamic forces at time t_n . The other matrices are kept constant,

since small volume element variations are assumed in relation to the small displacement assumption. In Section 4.3 a numerical study of the temporal integration schemes is presented.

2.4. Deformable interface tracking for contact detection

Collision detection mechanism is a non-trivial problem in transient contact mechanics. Generally based on the normal gap sign in Eq. (10), detection mechanisms require the calculation of the projection of the slave node onto the master surface. For a piecewise linear finite element surface, this projection may lead to existence or misdetection problems, especially for slave nodes situated at glottal exit, where the convexity of the vocal fold surface is noticeable and asymmetry of contacting surfaces may become apparent [14]. In vocal fold modeling, it is important to successfully detect contact at glottal exit, as collision forces acting in that specific area can be responsible for damages [4]. Furthermore, the choice of master faces and slave nodes cannot be made a priori due to the vocal fold geometry and a combinatorial search must be performed with the corresponding computational cost [17]. In vocal folds collision, the contact detection problem is therefore typically simplified by enforcing contact at the mid-sagittal plane [10, 4, 9]. However, it is important to avoid this assumption, since asymmetries in the vocal fold dynamics play a role in abnormal phonation [28]. In the present work, a new vocal fold collision detection mechanisms is implemented within the Deformable Simplicial Complex (DSC) framework [29]. The DSC method is a topology-adaptive method for tracking deformable interfaces represented as a set of triangular faces separating a tetrahedral unstructured mesh in different mediums. The detection mechanism makes use of the airway mesh, as will be described next. It allows for asymmetric collision and it supports both the penalty and the Lagrange multiplier formulations. For the sake of simplicity, the tissue mesh topology is fixed during the deformation, and only the vertices are advected.

The proposed contact detection mechanism is based on oriented volume inversion of an intraglottal conforming tetrahedral mesh, connecting the two sides of the vocal folds, as shown in Figure 2. Once a conforming airway tetrahedron is inverted, the nodes \mathbf{x}^s , $\mathbf{x}_{k=1,2,3}$ of a node-to-surface contact element can be defined, by making use of nodes labeling. As a consequence of the simplicity of the flow model, a criterion to stop the flow is defined. Hence, the flow is zero whenever any of the inverted airway tetrahedra has barycentric coordinates (x_{bc}, y_{bc}, z_{bc}) such that $|y_{bc}| \leq y_i \cdot \tau$, where τ is a zero flux parameter and y_i is the length of the major axis in Figure 1. By enforcing this condition, the Bernoulli airflow is assumed to be independent of variations in the anteroposterior dimension below a certain threshold defined by τ .

The constrained contact problem can be solved by the Lagrange multiplier method in Eqs. (25). In order to avoid projection existence problems that may arise at glottal exit, the domain in which the algorithm seeks $\bar{\mathbf{x}}^m$ can be expanded. For a given relaxation parameter δ with area units, the slave node projects onto the master surface if

$$\frac{1}{2} \left(\sum_{k=1,2,3} \sum_{l=1,2,3} \|(\mathbf{x}_k - \bar{\mathbf{x}}^m) \times (\mathbf{x}_l - \bar{\mathbf{x}}^m)\|_2 - 2 \cdot \|(\mathbf{x}_2 - \mathbf{x}_1) \times (\mathbf{x}_3 - \mathbf{x}_1)\|_2 \right) \leq \delta, \quad (27)$$

where $\|\cdot\|_2$ is the 2-norm distance. Since the condition is related to the geometry illustrated in Figure 2, the relaxation parameter is expected to depend on the tetrahedral mesh used for the computations. Therefore, a general value cannot be given, and only a trial and error approach can provide a good parameter.

As pointed out earlier, the penalty formulation in Eq. (22) can be simplified. Once a negative element volume V^e is found, the compressive reaction force on a colliding element may be approximated numerically as

$$-(\mathbf{g}^e)^T \kappa \mathbf{g}^e \cdot (\mathbf{x}^s \mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3)^T \approx (-\mathbf{n} \mathbf{n} \mathbf{n} \mathbf{n})^T \kappa \frac{V^e}{4}, \quad (28)$$

where $\mathbf{n} = \bar{\mathbf{n}}^m$ is the outward normal in Figure 2. Assembly may be performed to approximate the contact force vector $-\mathbf{G}^T \kappa \mathbf{G} \mathbf{x}$ in Eq. (22), to be added to the external force vector \mathbf{F} . The

Table I. Contact algorithm with the Lagrange multiplier method (II.3.A) and a penalty method (II.3.B). The penalty κ , relaxation δ , and zero flow τ parameters control the heuristics of the algorithm.

<pre> n ← 0 I. Initialization: find labels \mathcal{L} of nodes $(x_{i>0}, 0, z_i)$ and $(x_{i<0}, 0, z_i)$ build matrices \mathbf{M}, \mathbf{C}, \mathbf{K} and \mathbf{D} save material coordinates \mathbf{X} set initial velocity \mathbf{v}_0 collision_mid = false II. while n < number of time steps do II.1. compute aerodynamic forces, build \mathbf{F}_n: for all nodes in \mathcal{L} do compute areas for Eq. (3) find $a(z_{det}) = a_{min}$. compute volumetric flow rate: if collision_mid then stop flow else use Eq. (2) for all interface nodes do below z_{det}, find closest node in \mathcal{L}; use Eq. (17) above z_{det}, jet end for II.2. solve for current positions \mathbf{x}_n: collision_mid = collision = false set \mathbf{G}_{n+1} to zero solve Eq. (25) II.3. check for collision: for all intraglottal tetrahedra do if volume intraglottal tetrahedra $V^e < 0$ then collision = true </pre>	<pre> get collision data: define \mathbf{x}^s, $\mathbf{x}_{k=1,2,3}$ II.3.A. Lagrange method: if Eq. (27) then save \mathbf{x}^s, $\mathbf{x}_{k=1,2,3}$ II.3.B. Penalty method: compute contact force by Eq. (28) update \mathbf{F}_n compute barycentric coordinates (x_{bc}, y_{bc}, z_{bc}) if $y_{bc} \leq y_i \cdot \tau$ then collision_mid = true end if end for if collision re-solve for current positions \mathbf{x}_n: II.3.A. Lagrange method: build \mathbf{G}_{n+1} using Eq. (20) solve Eqs. (25) if $\lambda > 0$ remove contact element and rebuild \mathbf{G}_{n+1} resolve Eqs. (25) end if II.3.B. Penalty method: solve Eqs. (25) end if n ← n + 1 end while </pre>
--	--

algorithm safely avoids the classical master-slave paradigm, in which normal gap computations and projections are required. The contact algorithm is presented in Table I.

The collision approach is based in a coarse airway mesh that cannot present inner vertices. Therefore, an obvious drawback of the algorithm is the incompatibility with a numerical Navier-Stokes model of the flow.

2.5. Numerical Solver

Earlier in the paper, it was noted that the mechanical equilibrium of a system with position-based constraints described by the Lagrange function occurs at a saddle point. This means that the system matrix in Eq. (25) is indefinite. Consequently, gradient methods, which are effective to solve symmetric positive definite linear systems of equations, cannot be employed in the present model. As described in [29], the iterative generalized minimum residual method [30] is used in this study.

To improve the performance of the method, the preconditioning technique based on [31, 29] is well-suited for the present investigation. Hence, the system is preconditioned by multiplication of the inverse of a diagonal matrix

$$\mathbf{P} = \begin{pmatrix} \text{diag}(\mathbf{A}) & \mathbf{0} \\ \mathbf{0} & (1 + \alpha) h^2 \beta \cdot \text{diag}(\{ \mathbf{D}^T \mathbf{G}_{n+1}^T \}^T \cdot [\text{diag}(\mathbf{A})]^{-1} \{ \mathbf{D}^T \mathbf{G}_{n+1}^T \}) \end{pmatrix} \quad (29)$$

at both sides of Eq. (25), where $\text{diag}(\cdot)$ is a diagonal matrix with the diagonal elements in the argument, and $\{\cdot\}$ indicates an assembled matrix.

3. SIMULATION SETUP

The layered tissue geometry used in the simulations is adapted from [17], where elliptical shaped cross-sections are assumed, as stated earlier in Section 2.1; see Figure 1. The larynx is

Table II. Larynx geometry adapted from [17], parallel symmetric prephonatory intraglottal profile (in cm). Spatial discretization of the vocal fold boundary assumes 51 elliptical shaped cross-sections; see Figure 1 and Eq. (3). For the sake of brevity, some relevant sections are shown. $y_i = 0.7$ cm for all i , and cover and ligament has a thickness of 0.05 cm between section S_{17} and S_{27} .

	z_i	$ x_{i,<} = x_{i,>} $		z_i	$ x_{i,<} = x_{i,>} $
S_0	-2	0.7	\vdots		
S_7	-0.78	0.6	S_{26}	-0.04	0.025
S_8	-0.45	0.25	S_{27}	-0.02	-0.05
S_{12}	-0.42	0.16	S_{36}	0.03	0.3
S_{17}	-0.35	0.05	S_{43}	0.11	0.5
S_{18}	-0.31	0.025	S_{51}	2	0.7

Table III. Tissue parameters used in the finite element simulation for a normal larynx, adapted from [17]. (E, ν) , transversal Young's modulus and Poisson's ratio; (E', μ', ν') , longitudinal Young's and shear modulus, and Poisson's ratio. η indicates viscosity. All values are specified in the centimeter-gram-second system of units.

	E	ν	E'	μ'	ν'	η
Cover	$1 \cdot 10^4$	0.66	$10 \cdot 10^4$	$10 \cdot 10^4$	0.66	4
Ligament	$3 \cdot 10^4$	0.58	$20 \cdot 10^4$	$40 \cdot 10^4$	0.58	10
Body	$20 \cdot 10^4$	0.45	$40 \cdot 10^4$	$30 \cdot 10^4$	0.45	10

discretized into 51 cross-sections with constant major axis $y_i = 0.7$ cm and minor axis shown in Table II, and it is embedded into a cylinder of radius 0.8 cm and length 2 cm. The geometrical description of the vocal folds correspond to a parallel and symmetric prephonatory profile, a glottal gap of 0.05 cm, and vocal fold vertical thickness and lateral depth of 0.9 cm and 0.77 cm, respectively. No false vocal folds are included. For an average height of the intraglottal interface triangles of 0.02 cm, discretization of the narrow intraglottal airway in 9 sections (see S_8 to S_{26} in Table II) is chosen to keep low resolution Bernoulli flow while self-sustained oscillations are assured.

The tissue properties can be found in Table III. The values $\rho = 1.02$ g/cm³ and $\rho_{air} = 1.23 \cdot 10^{-3}$ g/cm³ are chosen for the tissue and air density, respectively. The driving subglottal pressure is set to $p_{sub} = 0.8 \cdot 10^4$ dyn/cm². The initial tissue velocity is set to zero.

The value $\tau = 0.5$ is chosen for the zero flow parameter, with the aim of not favoring extreme cases. The relaxation parameter for projection is set to $\delta = 0.2$ by trial and error, leading to a condition in Eq. (27) that successfully detects inverted tetrahedral at glottal exit for the tetrahedral mesh used in the simulations. The default temporal discretization scheme for calculations is the implicit, second-order accurate Hilbert-Hughes-Taylor α -method. The parameters employed are $\alpha = -0.3$ and time step increment $h = 50$ μ s.

The finite element mesh consists of 24584, 12754 and 17646 tetrahedra for body, ligament and cover, respectively. With regard to the labeling, the tissue layers are labeled differently for the left ($x < 0$) and right ($x > 0$) vocal folds. Furthermore, linear shape functions are used for the spatial finite element discretization. The relative tolerance of the iterative generalized minimum residual algorithm is 10^{-8} in all simulations.

4. RESULTS AND DISCUSSION

4.1. Model validation

Validation of the present vocal fold model is done by comparison of typical outputs with published numerical results in [5, 17, 6, 4, 10, 32] and experimental results in [33, 34]. In this

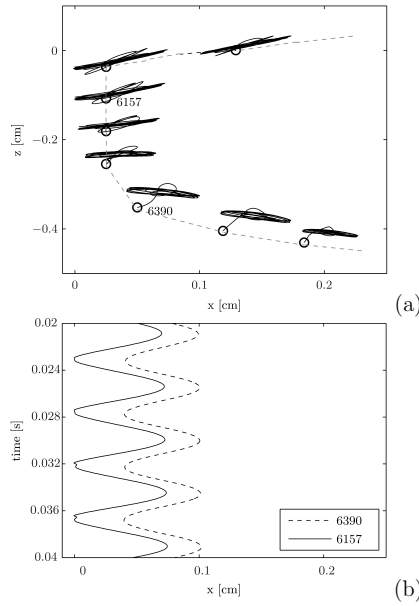


Figure 3. (a) Projected trajectories of selected interface nodes on the mid-coronal plane during the first 0.04 s. Circles corresponds to the initial configuration. Labels of two nodes, 6157 and 6390, with initial material coordinates (0.0250, 0, -0.1077) and (0.0500, 0, -0.3519), respectively, are shown. (b) Waveform of x coordinate (mediolateral direction). Solid line corresponds to node 6157, and dashed line to node 6390.

section, the contact model to be validated is solved by the Lagrange multiplier approach, and further validated by comparison with the commercial software ANSYS® Academic Research, Release 16.1.

The spatial coordinates of selected interface nodes placed on the mid-coronal plane are shown in Figure 3. The trajectories projected on a coronal plane are shown in Figure 3a, where circles indicate the initial configuration, making the initial transient apparent. The gray dashed line illustrates the initial superficial state of the vocal fold. The amplitude of oscillation in the vertical direction increases as the nodes approach the glottal exit, in agreement with the results in [17]. However, notice that the trajectories slightly differ from the paths measured on excised larynges [34]. In the present collision study, the choice of the vocal fold tissue parameters is made to simulate significant vocal fold contact to the detriment of highly realistic trajectories. Furthermore, Figure 3b shows the waveforms of the x spatial coordinates of nodes 6157 and 6390 in Figure 3a with initial material coordinates (0.0250, 0, -0.1077) and (0.0500, 0, -0.3519), respectively. The oscillations in the first 0.02 s are omitted for clarity reasons. The nodes present admissible theoretical amplitudes of oscillation [5]. The fundamental frequency of oscillation is about 219 Hz, which is an acceptable value for female phonation voices. Node 6157 placed in the narrow glottal region is affected by collision, which is apparent from the waveform near the mid-sagittal plane ($x = 0$). Compared to the results in [17], in the present model collision forces act to avoid body interpenetration. Figure 4 shows a coronal view of the glottal shape during a glottal cycle. The model has the ability to capture the alternate convergent-divergent shapes [35, 36]. A diverging shape can be observed during the closing phase (Figure 4a), followed by a converging one at collision (Figure 4b),

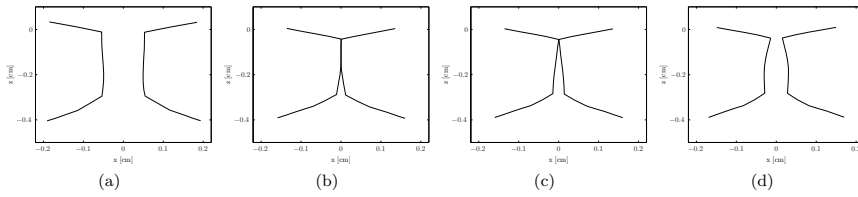


Figure 4. Vibration pattern for a subglottal pressure 0.8 kPa at times $t = 0.0353, 0.0366, 0.0367, 0.0375$ s from left to right, respectively.

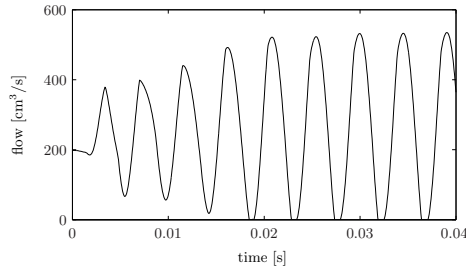


Figure 5. Volumetric flow rate U approximated as in Eq. (2). Approximate steady oscillatory behavior follows an initial transient of duration about 0.02 s.

which is maintained throughout the opening phase (Figures 4c-d). Notice in Figure 4b the significant contact region.

In Figure 5 the volumetric flow rate during transient and steady oscillations is shown. The maximum flow is between 500 and 600 cm^3/s . For a Bernoulli flow and similar initial cross-sectional areas [6], this is a feasible value. Furthermore, the open quotient in the simulations is about 0.87, slightly lower than the value reported in the computational model of [5]. Nevertheless, this value is above typical experimental data reported in the literature [32]. High open quotients have been shown to be related to high fundamental frequencies [32], and low subglottal pressures [17]. In the present model, a decrease of the subglottal pressure from 0.8 kPa to 0.6 kPa produces an increase of the open quotient from 0.87 to 0.89, and a frequency increase from 219 Hz to 222 Hz, in agreement with the results in [17]. Hence, with regard to the closing time, the present model is capable to simulate a similar trend as observed in previous studies. Finally, the open quotient highly depends on the solution method, as will be shown in the following section.

The performance of the proposed collision method is firstly examined by comparison to numerical solutions obtained with the commercial software ANSYS®. A simple impact problem is defined as two solid cylinders with the same dimensions and initial velocities 35 m/s, placed at a distance of 0.007 m; see Figure 6a. The material of the two cylinders is considered linear elastic with complete isotropy. The Young's modulus is $E = 2 \cdot 10^9$ and the Poisson's ratio $\nu = 0.3$. Frictionless contact is assumed. For comparison reasons, the contact problem in the commercial software is formulated in a Lagrangian form, and the contact searching algorithm is based on a master-slave configuration (target-contact in ANSYS terminology) with normal direction from the master surface, which corresponds to the cylinder at the left of Figure 6a. However, given the symmetry of the problem, the auto-symmetric behavior option in ANSYS® is activated. This means that the slave and master surface can be reversed locally. In the proposed collision model, the coarse conforming mesh for contact detection is generated in a region of possible contact defined a priori, in this case between the two cylinders in the cube $[0, 2.7] \times [-1, 1] \times [-1, 1]$. Notice that some boundary elements of the conforming mesh are

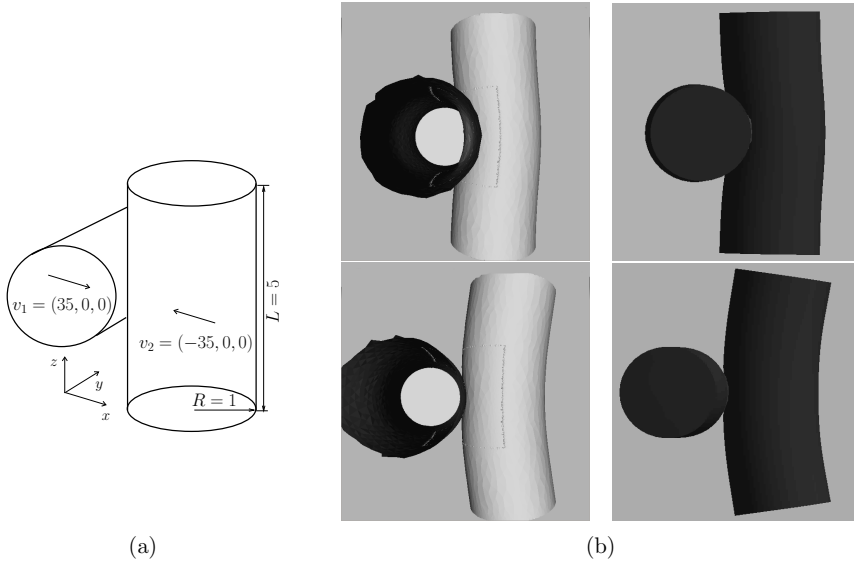


Figure 6. (a) Geometry of the impact problem. The origin is located at the center of the cylinder at the left of the illustration. All values in the SI of units. (b) Simulation of two colliding cylinders using the custom collision mechanism (left) and the commercial software ANSYS (right) at times $t = 0.0011$ s (top) and $t = 0.003$ s (bottom).

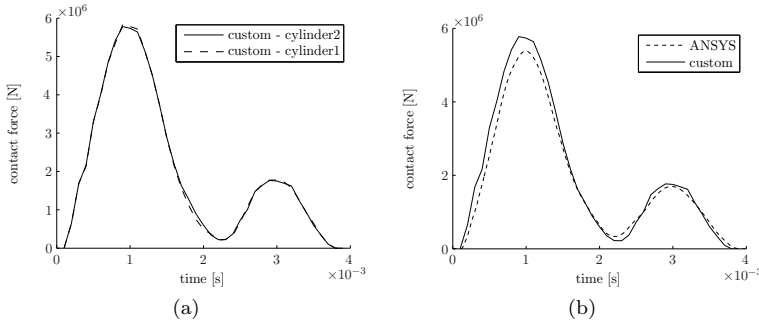


Figure 7. Magnitude of the contact force computed from the summation of all nodal contact forces applied to each of the cylinders (a); comparison to the contact force calculations obtained with the commercial software ANSYS (b).

partially visible in Figure 6b to illustrate the region of possible contact. Figure 6b shows the deformation of the colliding cylinders at times $t = 0.0011$ s (top) and $t = 0.003$ s (bottom) obtained with the proposed contact model (left) and ANSYS® (right). Only the boundary triangles are shown in the custom simulations for clarity reasons. The deformations are reasonable and show good agreement between the two simulations. Furthermore, the contact searching algorithm based on tetrahedral volume inversion seems to successfully detect contact. The selected times correspond to peak values of the total contact force, as Figure 7 shows. In Figure 7a the magnitude of the total contact force applied to each of the cylinders is shown.

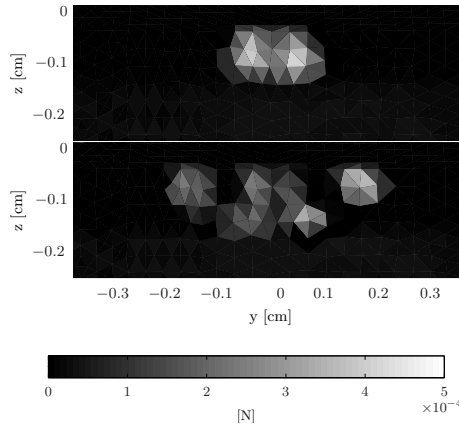


Figure 8. Sagittal view of the superficial layer of the right vocal fold ($x > 0$) during collision, from top to bottom at $t = 0.0371, 0.0372$ s. In grayscale, the magnitude of the x component of the contact and aerodynamic forces acting on interface triangles. Downstream the lowest contact node zero aerodynamic forces are considered due to the no vocal tract coupling and the flow stop criterion defined by the zero flux parameter $\tau = 0.5$ (see Section 2.4 and Table I).

The similarity between the two curves seems to be in agreement with the action-reaction principle. In Figure 7b the total contact force applied to one of the cylinders is compared to the results obtained with ANSYS®. The difference between the two simulations is between 2.8% and 6% for extreme values. This deviation may be explained by the relaxation of the projection condition in Eq. (27), which can lead to slight overestimation of the reaction forces. Furthermore, the results obtained with the proposed collision methodology show a compressive stress peak of about $4.89 \cdot 10^5$ kPa, which is 1.2% lower than the maximum compressive stress $4.95 \cdot 10^5$ kPa computed in ANSYS. Therefore, the proposed contact model provides numerical solutions similar to the results obtained with the cited commercial software for simple impact problems.

In regard to vocal fold collision, Figure 8 shows a sagittal view of the vocal folds and the magnitude of the x component of the contact and aerodynamic forces acting on the right vocal fold ($x > 0$) during vocal fold separation, at contact times $t = 0.0371, 0.0372$ s, from top to bottom. The contact force at each of the interface triangle is computed from the summation of contact forces applied to the vertices. Above the contact region the force is zero, since no vocal tract coupling is considered in this model. From the figure, the contact detection mechanism seems to successfully detect contact at glottal exit. In agreement with the studies presented in [10], maximum contact force occurs at the mid-coronal plane. With regard to the pressure distribution in the anteroposterior direction, for a minimum prephonatory glottal half widths of 0.025 cm between $z = -0.25$ and -0.03 cm, collision occurs along approximately 40% of the whole length of the vocal fold, which is similar to the results obtained in [10]. As in [4], values are of order of magnitude 10^{-4} in N. Finally, the relation between peak contact stress values and subglottal pressures is demonstrated in Table IV. The mediolateral component of the contact stress appears to increase as a function of the suglottal pressure, in agreement with the normal stress measurements in [33]. A simple linear regression reveals a slope of 1.72 and a coefficient of determination R^2 of 0.989. Furthermore, the estimated values belong to the interval [0.4 1.5] kPa. All these values are similar to the ones reported in [33].

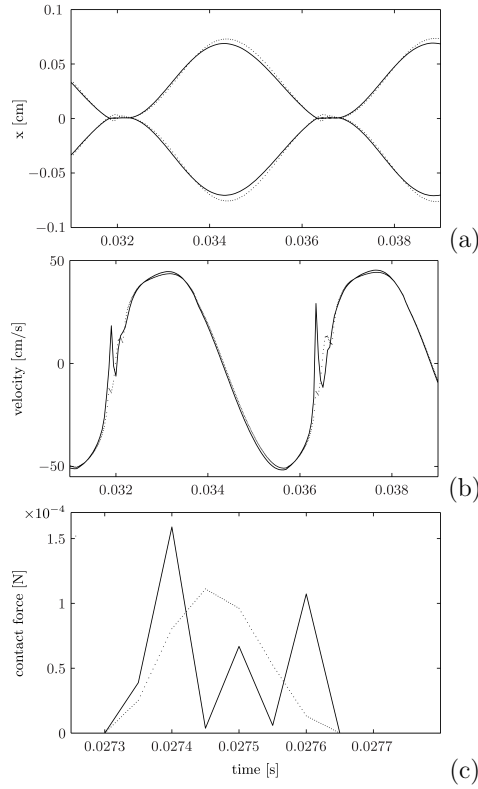


Figure 9. From top to bottom: x coordinate of spatial position, velocity, and nodal contact force for an interface node at initial material coordinates $(0.024, -0.136, -0.037)$. For comparison reasons, the top figure includes the spatial position of the opposite interface nodes at material coordinates $(-0.024, -0.136, -0.037)$. Solid lines indicates the results for the Lagrange formulation; dotted lines indicate the results for a penalty formulation with $\kappa = 10^7$.

4.2. Contact constraint enforcement

The vocal fold contact problem is here solved by the Lagrange multiplier method, and a penalty method with the approximation in Eq. (28), which takes advantage of the conforming airway mesh. The penalty method is applied to investigate the effect of inexact position-based contact constraint enforcement [9] and spring-like contact forces on the dynamics and mechanics at vocal fold contact. These results are compared to the proposed Lagrange multiplier solution for vocal fold contact.

Table IV. Peak mediolateral contact stress for different subglottal pressures. For comparison reason, units are in kPa [33].

contact stress [kPa]	0.54	1.17	1.21
p_{sub} [kPa]	0.6	0.8	1
	{slope, R^2 }		{1.72, 0.989}

Figure 9 shows the results obtained with the Lagrange multiplier method (solid line), and the results obtained with a penalty method with $\kappa = 10^7$ (dotted line). In Figure 9a waveforms of the x spatial coordinate of two opposite interface nodes placed at initial material coordinates $(0.024, -0.136, -0.037)$ and $(-0.024, -0.136, -0.037)$ are shown. The figure illustrates how the Lagrange multiplier solution is capable to simulate realistic displacements, while the penalty solution exhibits a noticeable volume interpenetration. In Figures 9b and 9c the x component of the tissue velocity and the nodal contact force applied to an interface node with initial position $(0.024, -0.136, -0.037)$ is shown, respectively. Note the spurious non-smooth behavior that arises at collision in the Lagrange multiplier solution. It may be argued that the abrupt changes in nodal contact force calculation are a consequence of the non-smoothness introduced by the contact conditions in Eq. (11). Regularization of the exact impact solution is common practice in numerical mechanics, which can be performed by several techniques. For instance, in [37] the Lagrange multipliers are computed at a time prior to collision instead, with a contact matrix built from the collision positions. Physically, the resulting smoothing effect may be explained by the compressibility of the air previous to collision modeled by the Lagrange multipliers. Nevertheless, an explicit scheme has to be used to temporally discretize the problem, as explained in [37]. These schemes are not unconditionally stable, which is not a desirable feature for complex mechanical problems as the present one. Alternatively, the solution may be regularized by means of other methods such as the Augmented Lagrange method [25], or Nitsche's method [38]. However, notice in Figure 9c that temporal integration of the two solutions give similar results. This observation suggests that the spatially integrated Lagrange contact force may show a different trend. This is supported by the results in Figure 10 that shows the mediolateral component of the impact stress computed with the Lagrange multipliers and the penalty methods, for a subglottal pressure of 0.8 kPa. The strong non-smooth behavior of the nodal contact force in Figure 9c is less apparent. In contrast to the penalty solution, the resulting Lagrange stress curve presents a sudden increase at first moment of impact, followed by a quick decrease. Then, the pressure decreases gradually, with a small positive increment which precedes the final contact phase. This behavior is in very good agreement with the impact pressure measurement results in [33], that show a sharp pressure pulse at the impact phase, and a pressure buildup in the preopen phase. This final impact pressure increase is also apparent in the results for the cylinders problem in Figure 7. The penalty solution, however, fails at reproducing the peak stress consequence of the high kinetic energy previous to collision; notice the high tissue velocities previous to impact in Figure 11b. This reveals the limitation of linear spring-like vocal fold contact forces to quickly dissipate the large kinetic energy at impact. Previous studies on collision forces and tissue stresses [4] have shown that large collision forces may be a cause of tissue damage. Hence, a penalty vocal fold contact model may lack relevant information related to damage, and discrete vocal fold models may consider the use of non-linear spring models of contact.

The impact stress in Figure 10 show a large vertical gap between the two curves. Comparison with the results in Figure 9c makes apparent that contact area calculations are significantly affected by the solution method. Furthermore, as mentioned earlier in this paper, accurate contact dynamics estimation, such as contact area, is of importance in inverse problem formulations [8]. In addition, contact areas and collision forces have been shown to be dependent on uncertain parameters, such as the subglottal pressure [4]. Therefore, robust and good estimates of the dynamical behavior at contact are necessary for voice pathology analysis. The numerical study that follows investigates the performance of different theoretical models of the vocal folds with regards to features estimation.

The importance of enforcement of the position-based collision constraint is illustrated in Figure 11. In Figure 11a the maximum contact area computed from the summation of triangular interface areas with vertex loaded by a collision force is shown; the horizontal axis corresponds to the interpenetration volume normalized to the maximum intersection volume when the effect of contact forces is neglected. Furthermore, the results are shown for a subglottal pressure of $p_{sub} = 0.8$ kPa in black marks, and $p_{sub} = 0.6$ kPa in white marks.

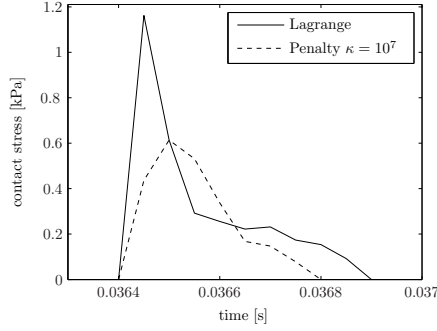


Figure 10. Mediolateral component of the contact stress with a subglottal pressure of 0.8 kPa for a Lagrangian and a penalty ($\kappa = 10^7$) solution.

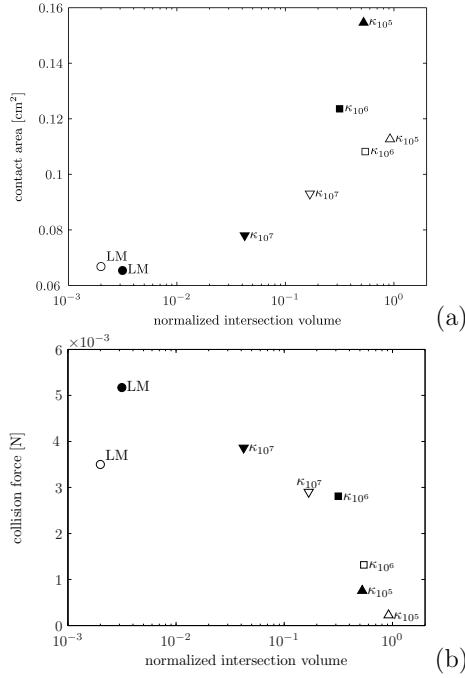


Figure 11. (a) Maximum contact area computed from the summation of interface triangles with any vertex loaded by a collision force. (b) Maximum x component of the total contact force. The horizontal axis corresponds to the intersection volume normalized to the maximum intersection volume when the effect of contact forces is neglected. Circles indicate results for the Lagrange multiplier method (LM); downward triangles, squares, and upward triangles correspond to a penalty formulation with $\kappa = 10^7$, 10^6 and 10^5 , respectively. Black and white marks indicate results for $p_{sub} = 0.8$ kPa and $p_{sub} = 0.6$ kPa, respectively. The selected collision time interval is $[0.036, 0.037]$, and the right vocal fold ($x > 0$) has been used for calculations.

The general increasing trend of the values suggests that overestimation of the contact area may occur as the non-penetration condition relaxes. In the same figure, circles correspond to the

results of the Lagrange multiplier method; downward triangles, squares, and upward triangles correspond to the penalty formulation with $\kappa = 10^7$, 10^6 and 10^5 , respectively. The maximum value of the penalty parameter shown in the results is constrained to the ill-conditioning of the system matrix and physical meaning of contact forces. The results show that the penalty parameter has a direct influence on contact area prediction. Furthermore, open quotients values of 0.85, 0.76 and 0.7 are predicted for the reported penalty parameters, respectively, with a subglottal pressure of $p_{sub} = 0.8$ kPa. This means that inexact position-based contact constraint enforcement has an effect on the temporal distribution of the transfer of kinetic energy at impact. Furthermore, in [17] it is argued that high kinetic energy is related to low open quotients. Inspection of Figure 11a reveals slight larger amplitudes of the penalty solution compared to the Lagrange multipliers solution, which gives an open quotient of 0.87. Hence, increasing interpenetration leads to a spurious increase of the kinetic energy and contact areas, and decreasing open quotients, in agreement with [17]. Finally, note that small subglottal variations do not have a strong effect on contact area estimations, as large contact area variations in the penalty solution may be explained by spurious kinetic energy variations. Therefore, the contact area is dependent to subglottal pressures, as predicted in [4], but also highly dependent to the contact constraint enforcement.

In regard to contact force estimation, Figure 11b shows the maximum x component of the total contact force. The observed general trend indicates an effect of the vocal fold interpenetration, as expected from Eq. (28); hence, the value of the penalty parameter is crucial. Furthermore, comparison of collision forces with $p_{sub} = 0.8$ kPa (black marks) and $p_{sub} = 0.6$ kPa (white marks), shows that contact forces decrease with subglottal pressure, regardless of the method used to enforce the contact constraint. Finally, note that the peak impact stress in Figure 10 does not correspond to the ratio between maximum total contact force and maximum contact area in Figure 11 as peak values do not occur at the same time, in agreement with [4].

Due to the difficulties in extracting experimental model parameters, arbitrary choice of the input contact parameters is generally made [6]. Hence, the number of unknown input parameters introduces uncertainty in the model output, as demonstrated earlier. In this sense, the Lagrange multiplier method appears to be a solution to reduce model uncertainties by reducing the number of unknowns and introducing robustness. Finally, notice that the solution obtained by the Lagrange multiplier method reduces the non-physical interpenetration volume by a factor of 10^3 , but still it is not zero as one may expect from the theory. This can be explained as a consequence of numerical errors and the relaxed definition of slave node projection in Eq. (27).

In summary, inexact contact constraint enforcement has a clear effect on vocal fold dynamics. The proposed Lagrange vocal fold contact solution is capable to successfully replicate the temporal distribution of the impact stress in contrast to linear spring-like contact forces. Furthermore, it is able to introduce robustness and reduce uncertainties in the model output. Nevertheless, the Lagrange approach may lead to under and overestimation of the instantaneous nodal contact force and the tissue velocity at contact.

4.3. Temporal integration

Earlier in Section 2.3.2, the importance of temporal integration schemes with respect to system dynamics was emphasized. A numerical study of a number of schemes that are unconditionally stable and second-order accurate under certain of premises (see Section 2.3.2) is here presented. Due to the complexity of the system, methods with stability conditions such as the α -method with $\alpha = -0.41421$ used in [39] are not considered.

In regard to nodal force collision calculations, the temporal integration scheme appears to have an influence as Figure 12 shows. In the figure, circles correspond to the undamped Newmark scheme in Eq. (24) with $\gamma = \frac{1}{2}$ and $\beta = \frac{1}{4}$, similar to e.g., [9]; triangles indicate the α -method with $\alpha = -0.3$; solid and dotted lines indicate the results with the Lagrange multiplier and the penalty methods ($\kappa = 10^7$), respectively. The spurious non-smooth behavior worsens

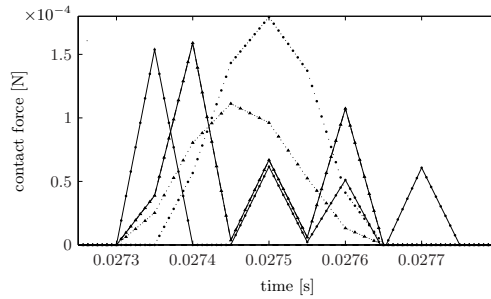


Figure 12. Contact force of a interface node at initial material coordinates $(0.024, -0.136, -0.037)$, as in Figure 9c. Circles correspond to an undamped Newmark with $\gamma = \frac{1}{2}$, and triangles to the α -method with $\alpha = -0.3$. Dotted line indicates a penalty method with $\kappa = 10^7$, and the solid line, the Lagrange multiplier method, as in Figure 9. Time step is $h = 50 \mu\text{s}$.

for an undamped Newmark method. It might be argued that the lack of negative numerical damping leads to an increase of inertial forces and large amplitude of oscillation, and thereby a worsening of the non-smoothness. As expected, large amplitudes of vocal fold oscillation have been observed with the undamped temporal scheme, although the results are not shown here for the sake of brevity. Notice that the increase of amplitude has a clear effect on the estimation of the penalty nodal contact force solution, in agreement with the results in Figure 11. Furthermore, no noticeable phase error can be seen from the temporal discretization simulations. Finally, the numerical solutions have a stable behavior with a time step of $50 \mu\text{s}$, which supports the unconditional stability condition.

In summary, the results suggest that α -methods are superior to Newmark schemes as numerical damping appears to be advantageous in vocal fold contact simulations.

4.4. Asymmetric collision study

Asymmetric vocal fold oscillations are typically observed in several voice pathologies such as mass lesions, intracordal scarring or vocal fold paralysis [28]. This has motivated numerical studies on the effect of abnormal oscillations on phonation [12, 15]. In contrast to previous studies on collision-free asymmetric oscillations [12], the proposed contact detection algorithm allows for the study of the effect of asymmetric collision on vocal fold mechanics. Similarly to [12, 15], the present work limits to asymmetric vibration induced by a tension imbalance between left and right cover layer. Thus, structural asymmetries typical of mass lesions are not included, which would possibly require the modeling of frictional contact. All simulations use a constant subglottal pressure of 0.8 kPa .

A tension mismatch $100 \cdot (E_{\text{right}} - E_{\text{left}})/E_{\text{right}}$ of 20% and 40% between left and right longitudinal Young's modulus of the cover layer is enforced. This means that the right vocal fold is stiffer than the left vocal fold. Figure 13 compares the left and right vocal fold mechanical response averaged over 4 oscillatory cycles immediately after the initial transient response (see Figure 5). The results with symmetric mechanical properties are shown for comparison reasons. All values correspond to the arithmetic mean. To illustrate the asymmetric oscillations, solid horizontal lines indicate the x coordinate of the collision point of node 6157, normalized to the x coordinate of its initial material coordinate for clarity reasons. From the results, the collision point appears to be successfully detected by the contact detection algorithm. Furthermore, black dot clips correspond to the ratio between peak-to-peak x displacement of node 6157 and the opposite node laying on the left vocal fold; this is equivalent to a classical measure of asymmetry in clinical research [28]. In the figure, the ratio seems to increase with the tension mismatch, indicating increasing asymmetries in the amplitude of vibration. In dashed lines, the collision-free solution is shown by neglecting the effect of surface contact forces. These

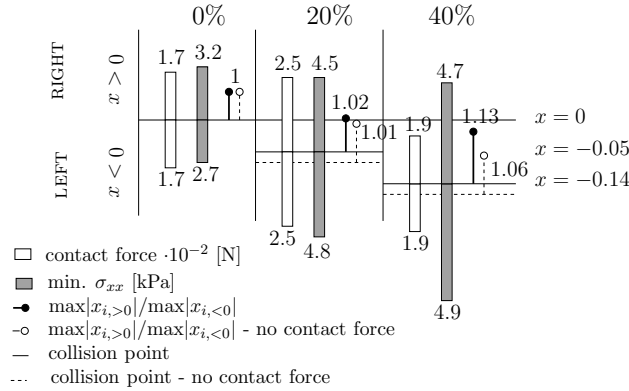


Figure 13. For each vocal fold, total contact force in N (white bar), maximum compressive stress in kPa (gray bar) and ratio between right and left vocal fold vibration amplitude (dot clip) for 0%, 20% and 40% longitudinal Young's modulus mismatch of the cover layer. Solid horizontal line indicates the x spatial coordinate of the collision point of the node 6157, normalized to the initial material coordinate. Dashed lines indicate that the effect of surface contact forces is neglected. Values are obtained from the average over 4 cycles after transient oscillations.

results indicate that the collision forces, and thereby the energy distribution, play a role on asymmetric vocal fold self-oscillations. In connection with the results in Figure 11b, inexact contact constraint enforcement, widely accepted in the vocal fold modeling community, has an effect on asymmetric dynamics.

With regard to internal and reaction forces, gray and white bars in Figure 13 show the maximum compressive stress and the x component of the Lagrange solution of the contact force integrated over a collision interval, respectively. As the stiffness asymmetry increases, an increasing compressive stress is apparent. It can be hypothesized that asymmetric oscillations increase the risk of damage since large compressive stress may be related to vocal fold damage [3]. Nevertheless, impact forces do not appear to be correlated with the degree of asymmetry. Previous studies have related mass lesions with impact forces at vocal fold collision [40]. Hence, from the present results asymmetric collision can not be considered a sufficient factor for increasing risk of damage.

Furthermore, note in Figure 13 that the same mediolateral total contact force is applied to each vocal fold, and similarly, the small anteroposterior component, which is of order 10^{-6} N. This contradicts a number of results in [15], where asymmetry in the vocal fold oscillation is related to asymmetry in the anteroposterior component of the contact force. However, the action-reaction principle predicts equal forces, and hence, the present simulations validate the contact model proposed in this work. Finally, it is noteworthy that variances increase with tension imbalance, which may indicate unstable oscillation, as investigated in [6].

5. CONCLUSIONS

The asymmetric collision problem of a three-dimensional continuum model of the vocal folds driven by Bernoulli aerodynamic forces has been formulated in terms of position-based inequality constraints. In contrast to existing symmetric vocal fold contact models [17, 10, 15], the present contact model has been developed within a finite-element contact mechanics framework, providing mathematical details and a numerical analysis. This work covered the

entire contact modeling process, which includes definition of contact boundary condition, contact detection methodology, validation and numerical analysis.

A novel contact detection mechanism has been developed to solve the problems of symmetric contacting bodies, non-smooth boundaries, and asymmetric contact detection posed in [9] for the general case of non-matching tissue meshes. Firstly, the algorithm safely eliminates the need of predefinition of the master and the slave bodies, which poses a problem for vocal folds with a symmetric geometry. Secondly, it deals with the projection point problem at convex boundaries such as the upper vocal fold surfaces. Hence, the algorithm successfully detects contact at glottal exit where collision forces may be responsible for tissue damage [4]. Thirdly, the models successfully detects asymmetric contact induced by tissue property asymmetries. However, the model may show limitations in voice pathologies investigations related to geometrical irregularities as frictional contact is expected to play a role. Nevertheless, the proposed node-to-surface contact model with the novel contact detection mechanism is extensible to frictional contact with relative ease. Therefore, future developments of the model will be devoted to vocal fold contact in asymmetric geometries.

Numerical analysis has been included in this study to assure a good performance of the vocal fold model. The effect of inexact collision constraint enforcement [9, 15, 11] and temporal integration schemes on vocal fold mechanics and dynamics has been investigated, the results showing that vocal fold contact is highly sensitive to non-physical energy variations. Different variational methodologies for constrained minimization problems have been applied to the vocal fold contact problem. It has been shown that linear spring-based contact models poorly replicate the time evolution impact phase observed in vocal fold measurements [33]. Furthermore, the penalty solution is highly sensitive to subglottal pressure variations and model parameter-choice. In an effort to reduce uncertainties in the model output and spurious energy variations, a vocal fold contact model based on the Lagrange multiplier method has been proposed. To reduce the computational cost of the Lagrange solution, a preconditioner to increase the convergence ratio of the iterative solver has been designed. The solution showed realistic contact stress simulations, robustness in terms of subglottal pressure variations, and accuracy in vocal fold contact dynamics estimation. However, the method is not suitable for predicting instantaneous nodal contact forces, which showed a spurious non-smooth behavior. Finally, the non-smooth vocal fold collision problem benefits from highly dissipative temporal discretization schemes. A Hilber-Hughes-Taylor α -method with $\alpha = -0.3$ seems to provide satisfactory results.

The present contact model is capable to predict specific mechanical behavior at asymmetric contact induced by stiffness mismatch enforcement, in contrast to the collision-free study in [12]. In contrast to the trend observed in impact force estimations, an increasing maximum compressive stress seems to correlate with the degree of asymmetry, which may contribute to additional damage.

Finally, the present finite-element vocal fold contact model may be adapted to advanced fluid-structure models with simplified symmetric contact (e.g., see [9, 11, 2]) to provide more insights into the effect of collision on phonation.

ACKNOWLEDGEMENT

This work was supported by grant number 110230 from the Swedish organization AFA Försäkring.

REFERENCES

1. Van Den Berg J, Zantema JT, Doornenbal RJ. On the Air Resistance and the Bernoulli Effect of the Human Larynx. *J. Acoust. Soc. Am.* 1957; **29**(5):626–631.
2. Šidlof P, Zörner S, Hüppe A. A hybrid approach to the computational aeroacoustics of human voice production. *Biomech. Model Mechanobiol.* 2015; **14**(3):473–488.
3. Titze IR, Svec JG, Popolo PS. Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues. *J. Speech Lang. Hear. R.* 2003; **46**(4):919–932.

4. Gunter HE. A mechanical model of vocal-fold collision with high spatial and temporal resolution. *J. Acoust. Soc. Am.* 2003; **113**(2):994–1000.
5. Alipour F, Berry DA, Titze IR. A finite-element model of vocal-fold vibration. *J. Acoust. Soc. Am.* 2000; **108**(6):3003–3012.
6. Steinecke I, Herzel H. Bifurcations in an asymmetric vocal-fold model. *J. Acoust. Soc. Am.* 1995; **97**(3):1874–84.
7. Horaček J, Sidlof P, Švec JG. Numerical simulation of self-oscillations of human vocal folds with Hertz model of impact forces. *J. Fluid. Struct.* 2005; **20**(6):853–869.
8. Spencer M, Siegmund T, Mongeau L. Determination of superior surface strains and stresses, and vocal fold contact pressure in a synthetic larynx model using digital image correlation. *J. Acoust. Soc. Am.* Feb 2008; **123**(2):1089–1103.
9. Zheng X, Xue Q, Mittal R, Bielamowicz S. A coupled sharp-interface immersed boundary-finite-element method for flow-structure interaction with application to human phonation. *J. Biomech. Eng.* 2010; **132**(11):111003 (1–12).
10. Tao C, Jiang JJ, Zhang Y. Simulation of vocal fold impact pressures with a self-oscillating finite-element model. *J. Acoust. Soc. Am.* 2006; **119**(6):3987–3994.
11. Bhattacharya P, Siegmund TH. Computational modeling of vibration-induced systemic hydration of vocal folds over a range of phonation conditions. *Int. J. Numer. Meth. Biomed. Engng.* 2014; **48**:169–187.
12. Zhang Z, Hieu Luu T. Asymmetric vibration in a two-layer vocal fold model with left-right stiffness asymmetry: Experiment and simulation. *J. Acoust. Soc. Am.* 2012; **132**(3):1626–1635.
13. Xue Q, Mittal R, Zheng X, Bielamowicz S. A computational study of the effect of vocal-fold asymmetry on phonation. *J. Acoust. Soc. Am.* 2010; **128**(2):818–827.
14. Simo J, Laursen T. An augmented lagrangian treatment of contact problems involving friction. *Comput. Struct.* 1992; **42**(1):97–116.
15. Xue Q, Mittal R, Zheng X, Bielamowicz S. Computational modeling of phonatory dynamics in a tubular three-dimensional model of the human larynx. *J. Acoust. Soc. Am.* 2012; **132**(3):1602–1613.
16. Wriggers P. *Computational contact mechanics*. John Wiley & Sons, 2002.
17. Rosa MDO, Pereira JC, Grellet M, Alwan A. A contribution to simulating a three-dimensional larynx model using the finite element method. *J. Acoust. Soc. Am.* 2003; **114**(5):2893–2905.
18. Sommer DE, Erath BD, Zanartu M, Peterson SD. Corrected contact dynamics for the Steinecke and Herzel asymmetric two-mass model of the vocal folds. *J. Acoust. Soc. Am.* 2012; **132**(4):EL271–EL276.
19. Story BH, Titze IR. Voice simulation with a body-cover model of the vocal folds. *J. Acoust. Soc. Am.* 1995; **97**(2):1249–1260.
20. Ishizaka K, Matsudaira K. *Fluid mechanical considerations of vocal cord vibration*. Monogr. 8, Speech Commun. Res. Lab.: Santa Barbara, CA., 1972.
21. Pelorson X, Hirschberg A, Van Hassel RR, Wijnands APJ, Auregan Y. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *J. Acoust. Soc. Am.* 1994; **96**(6):3416–3431.
22. Berry DA, Titze IR. Normal modes in a continuum model of vocal fold tissues. *J. Acoust. Soc. Am.* 1996; **100**(5):3345–3354.
23. Hughes TJR. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover, 2000.
24. Zienkiewicz O. *The finite element method*. McGraw-Hill, 1977.
25. Bertsekas D. *Constrained optimization and Lagrange multiplier methods*. Academic Press, 1982.
26. Newmark NM. A Method of Computation for Structural Dynamics. *J. Eng. Mech. Div.* 1959; **85**:67–94.
27. Hilber HM, Hughes TJR, Taylor RL. Improved Numerical Dissipation for time Integration Algorithms in Structural Dynamics. *Earthquake Eng. Struct. Dyn.* 1977; **5**:283–292.
28. Bless D, Hirano M, Feder RJ. Videostroboscopic evaluation of the larynx. *Ear, Nose & Throat Journal* 1987; **66**:289–296.
29. Misztal MK, Erleben K, Bargteil A, Fursund J, Christensen BB, Andreas Bærentzen J, Bridson R. Multiphase flow of immiscible fluids on unstructured moving meshes. *IEEE Trans. Vis. Comput. Graphics* 2014; **20**(1):4–16.
30. Saad Y, Schultz MH. GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. *SIAM J. Sci. Comput.* 1986; **7**(3):856–869.
31. Phoon KK, Toh KC, Chan SH, Lee FH. An efficient diagonal preconditioner for finite element solution of Biot's consolidation equations. *Int. J. Numer. Meth. Eng.* 2002; **55**(4):377–400.
32. Lohscheller J, Švec J, Döllinger M. Vocal fold vibration amplitude, open quotient, speed quotient and their variability along glottal length: kymographic data from normal subjects. *Logoped. Phoniater. Vocol.* 2013; **38**(4):182–92.
33. Jiang JJ, Titze IR. Measurement of vocal fold intraglottal pressure and impact stress. *J. Voice* 1994; **8**(2):132–144.
34. Döllinger M, Berry DA. Visualization and quantification of the medial surface dynamics of an excised human vocal fold during phonation. *J. Voice* 2006; **20**(3):401–413.
35. Titze IR. *The Myoelastic Aerodynamic Theory of Phonation*. National Center for Voice and Speech, 2006.
36. Zheng X, Mittal R, Xue Q, Bielamowicz S. Direct-numerical simulation of the glottal jet and vocal-fold dynamics in a three-dimensional laryngeal model. *J. Acoust. Soc. Am.* 2011; **130**(1):404–415.
37. Carpenter NJ, Taylor RL, Katona MG. Lagrange constraints for transient finite element surface contact. *Int. J. Numer. Meth. Eng.* 1991; **32**(1):103–128.
38. Nitsche J. Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh. Math. Semin. Univ. Hambg.* 1970; **36**(1):9–15.

39. Bhattacharya P, Siegmund T. A computational study of systemic hydration in vocal fold collision. *Comput. Methods Biomech. Biomed. Engin.* 2013; **17**(16):1835–52.
40. Titze IR. Mechanical stress in phonation. *J. Voice* 1994; **8**(2):99–105.

Paper B

Vocal fold collision modeling

Authors: Alba Granados, Jonas Brunskog, Marek Krzysztof Misztal

Published in: *Proceedings of the 9th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, pp.5-8, Firenze, Italy, 2015.

Errata: The total contact force estimations in Fig. 3 are overestimated. Compare to the correct results in Fig. 11(b) in Paper A. This source code error implementation does not affect local contact force and contact area calculations.

VOCAL FOLD COLLISION MODELING

A. Granados¹, J. Brunskog¹, M. K. Misztal²

¹ Acoustic Technology, Technical University of Denmark, Kongens Lyngby DK-2800, Denmark

² Niels Bohr Institute, University of Copenhagen, Copenhagen DK-2100, Denmark
algra@elektro.dtu.dk

Abstract: When vocal folds vibrate at normal speaking frequencies, collisions occurs. The numerics and formulations behind a position-based continuum model of contact is an active field of research in the contact mechanics community. In this paper, a frictionless three-dimensional finite element model of the vocal fold collision is proposed, which incorporates different procedures used in contact mechanics and mathematical optimization theories. The penalty approach and the Lagrange multiplier method are investigated. The contact force solution obtained by the penalty formulation is highly dependent on the penalty parameter value. Furthermore, the Lagrange approach shows poor results with regard to instantaneous contact force estimation. This motivates the use of an Augmented Lagrange approach to regularize the Lagrange contact force solution. Finally, the effect of the interpenetration volume on contact force and contact area computations is illustrated.

Keywords : Vocal folds, collision, constrained optimization, finite element method, contact detection.

I. INTRODUCTION

Mathematical descriptions of self-oscillating finite element models of the vocal folds have been reported in the literature (e.g., see [1]). A continuum model of the airflow coupled to a deformable three-dimensional body have been one of the main focuses. For purpose of clinical research, investigations on the mechanical conditions that arise during phonation are of special interest. At normal speaking frequencies, vocal fold collision occurs, and the tissue is affected by specific stresses and reaction forces [2]. Hence, a detailed mathematical study of the collision process is expected to contribute to a better understanding of vocal fold mechanics.

In the context of continuum mechanics, the vocal fold contact can be modeled by enforcing position-based constraints to the minimization of the total

potential energy of the mechanical system. Methodologies from mathematical optimization theory can be applied in order to solve the contact constrained problem [3]. In this paper, a Penalty method and a Lagrange multiplier approach are investigated for the case of frictionless vocal fold collision. Furthermore, a penalty regularization of the Lagrange multiplier method is carried out by the Augmented Lagrange technique applied together with an Uzawa type algorithm [3]. Finite element contact discretization and contact detection mechanism that allows for asymmetric collision are presented.

II. METHODOLOGY

A three-dimensional deformable viscoelastic model of the vocal folds driven by a Bernoulli glottal airflow is described. At each time step the new equilibrium position is found as the minimum of the total potential energy by the variational formulation. When collision occurs, the contact constrained minimization problem is solved by different methods.

A. Governing equations

In the absence of volume forces, the vocal fold deformation is described by the equation of balance

$$\nabla \cdot \boldsymbol{\sigma} = \rho \frac{\partial^2 \mathbf{x}}{\partial t^2} \quad \mathbf{x} \in \mathcal{V}_{\text{solid}}, \quad (1)$$

in the deformed state $\mathcal{V}_{\text{solid}} \subset \mathbb{R}^3$, the constitutive equation for a transversely isotropic linear viscoelastic solid as in [4], and the Dirichlet and Neumann boundary conditions

$$\mathbf{x} - \mathbf{X} = 0 \quad \text{in } \Gamma_D \subset \partial \mathcal{V}_{\text{solid}} \quad (2.1)$$

$$\boldsymbol{\sigma} \cdot \mathbf{n} = \mathbf{p} \quad \text{on } \Gamma_N \subset \partial \mathcal{V}_{\text{solid}} \quad (2.2)$$

respectively, where $\boldsymbol{\sigma}$ is the stress tensor, \mathbf{X} represents the material coordinates, \mathbf{n} is the outward normal, and \mathbf{p} is the aerodynamic pressure derived from Bernoulli's principle. The Dirichlet boundary where the displacement is zero is placed in the anteroposterior glottal regions; see [3] for further details. The equilibrium position can be found as the minimum of the total potential energy Π . Hence, for

admissible displacement variations or test functions \mathbf{w} that vanish in the Dirichlet boundary, the weak formulation of the problem takes the form

$$\delta \Pi = \int_{V_{solid}} \rho \mathbf{w}^t \cdot \frac{\partial^2 \mathbf{x}}{\partial t^2} dV + \int_{V_{solid}} \nabla \mathbf{w}^t : \boldsymbol{\sigma} dV - \int_{\Gamma_N} \mathbf{w}^t \cdot \mathbf{p} d\Gamma = 0, \quad (3)$$

where $\delta \Pi$ indicates the variation of the energy.

When collision between the vocal folds occurs, additional constraints may be activated on the contact area $\Gamma_C \subset \partial V_{solid}$. In order to avoid unphysical body interpenetration for a frictionless contact, non-negativeness of the normal gap between a superficial slave node \mathbf{x}^s and a master surface placed at the opposite vocal fold may be enforced by the position-based constraint

$$g_N = (\mathbf{x}^s - \bar{\mathbf{x}}^m) \cdot \bar{\mathbf{n}}^m \geq 0 \quad \text{on } \Gamma_C \subset \partial V_{solid}, \quad (4)$$

where $\bar{\mathbf{x}}$ is the projection of the slave node onto the master surface; see Fig. 1.

B. Contact constraint enforcement

A penalty, a Lagrangian and an augmented Lagrangian [3] methods are here studied to enforce the inequality constraint in Eq. (4). Only the Lagrangian solution enforces the collision constraint in exact form.

The penalty method consists of a minimization problem where the objective function involves the collision-free potential energy and a term which penalizes infeasible positions on Γ_C as

$$\Pi + \int_{\Gamma_C} \frac{1}{2} \kappa \|g_N(\mathbf{x})\|^2 d\Gamma, \quad (5)$$

where $\kappa > 0$ is a penalty parameter. Optimality conditions lead to the variational formulation

$$\delta \Pi + \int_{\Gamma_C} \kappa (\mathbf{w}^s - \bar{\mathbf{w}}^m) \cdot (\mathbf{x}^s - \bar{\mathbf{x}}^m) d\Gamma = 0 \quad (6)$$

to be combined with Eq. (3). The second term above can be interpreted as minus the reaction force to avoid interpenetration. For non-adhesion contact, the reaction force must be compressive. Hence, it can be seen that as the penalty parameter tends to infinity the normal gap tends to zero, and the optimal of the new minimization problem approaches the exact equilibrium solution at collision. However, large penalty parameters may lead to ill-conditioning of the global matrix.

The Lagrangian method solves the inequality constrained problem by solving the optimization problem with objective function

$$L(\mathbf{x}, \boldsymbol{\Lambda}) = \Pi + \int_{\Gamma_C} \boldsymbol{\Lambda} g_N(\mathbf{x}) d\Gamma, \quad (7)$$

called the Lagrangian function, where $\boldsymbol{\Lambda}$ is the Lagrange multiplier vector, also called dual variables.

Optimality conditions to the problem are

$$\delta \Pi + \int_{\Gamma_C} \boldsymbol{\Lambda} (\mathbf{w}^s - \bar{\mathbf{w}}^m) \cdot \bar{\mathbf{n}}^m d\Gamma = 0$$

$$\boldsymbol{\Lambda} \leq 0$$

$$\boldsymbol{\Lambda} \cdot g_N(\mathbf{x}) = 0, \quad (8)$$

which are known as the Karush-Kuhn-Tucker conditions for optimality. Note that the Lagrange multiplier vector can be seen as the compressive reaction forces. From a physical point of view, the last condition indicates that no contact forces are active when the normal gap is positive, and the non-penetration constraint is fulfilled in exact form whenever collision occurs. However, this approach introduces additional unknowns in the form of Lagrange multipliers. Furthermore, the Lagrangian approach in Eq. (8) is a non-smooth contact formulation, and regularization techniques may be used to improve results.

The Augmented Lagrange formulation combines the Lagrange and the penalty approaches, without additional unknowns. A simplified version is the Uzawa algorithm [3] which may be summarized as follows. For an initial Lagrange multiplier vector $\boldsymbol{\Lambda}_k$, a new equilibrium is found by minimization of

$$\Pi + \int_{\Gamma_C} \boldsymbol{\Lambda}_k g_N(\mathbf{x}) d\Gamma + \int_{\Gamma_C} \frac{1}{2} \kappa \|g_N(\mathbf{x})\|^2 d\Gamma, \quad (9)$$

where the last penalty term can be seen as a regularization term for non-smoothness. The Lagrange multiplier vector is updated in an augmentation iteration as

$$\boldsymbol{\Lambda}_{k+1} = \boldsymbol{\Lambda}_k + \min \{ \kappa g_N(\mathbf{x}_{k+1}), \boldsymbol{\Lambda}_k \}, \quad (10)$$

where \mathbf{x}_{k+1} is the solution of the minimization problem. The update in Eq. (10) can be seen as a gradient ascent algorithm, as the critical point of the Lagrangian in Eq. (7) occurs at a maximum over the multipliers [5]. As the contact constraint is not solved in an exact form, the augmentation procedure in Eq. (10) continues until a convergence criterion for $g_N(\mathbf{x}_{k+1})$ is fulfilled. Furthermore, the penalty parameter can be increased at each augmentation step to speed up the convergence rate. However, to avoid ill-conditioning of the system matrix due to a large

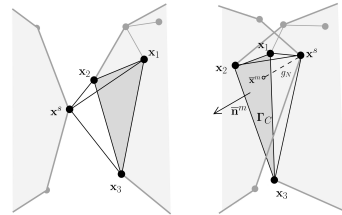


FIG. 1. Conforming interface mesh for collision detection; before (left) and after collision (right).

penalty parameter value, a maximum number of augmentations must be requested.

C. Spatial and temporal discretization

The spatial finite element discretization is based on a tetrahedral mesh. Hence, the interface domain is formed by triangular elements. As Fig. 1 illustrates, a coarse conforming interface tetrahedral mesh may be defined to detect contact; whenever an oriented interface element volume is inverted, collision occurs. Once the slave node \mathbf{x}^s and the master surface with vertices \mathbf{x}_1 , \mathbf{x}_2 and \mathbf{x}_3 are detected, by means of an isoparametric transformation with linear basis functions $N_i(\xi, \zeta)$ defined on a reference triangular element, the projection \mathbf{x}^m corresponds to the local coordinates $(\bar{\xi}, \bar{\zeta})$, and a contact element matrix

$$\mathbf{g}^e = (\bar{\mathbf{n}}^m - N_1(\bar{\xi}, \bar{\zeta})\bar{\mathbf{n}}^m - N_2(\bar{\xi}, \bar{\zeta})\bar{\mathbf{n}}^m - N_3(\bar{\xi}, \bar{\zeta})\bar{\mathbf{n}}^m) \quad (11)$$

with $\mathbf{g}^e \cdot (\mathbf{x}^s - \mathbf{x}_1 - \mathbf{x}_2 - \mathbf{x}_3) \geq 0$ contributes to the assembled global constraint contact matrix \mathbf{G} .

The penalty approach in Eq. (6) can be simplified further in the way that follows. Once a negative oriented element volume V^e is found, the compressive reaction force on a colliding element may be approximated numerically as

$$-(\mathbf{g}^e)^T \kappa \mathbf{g}^e \cdot (\mathbf{x}^s - \mathbf{x}_1 - \mathbf{x}_2 - \mathbf{x}_3) \approx \bar{\mathbf{n}}^m (-1 \ 1 \ 1 \ 1)^T \frac{\kappa V^e}{4} \quad (12)$$

Hence, a matrix \mathbf{F}_c can be assembled. For global mass, damping, and stiffness matrices \mathbf{M} , \mathbf{C} , and \mathbf{K} , respectively, and \mathbf{F} a vector of applied aerodynamic forces, the finite element system of the penalty approach is

$$\mathbf{M} \ddot{\mathbf{x}} + \mathbf{C} \dot{\mathbf{x}} + \mathbf{K}(\mathbf{x} - \mathbf{X}) = \mathbf{F} + \mathbf{F}_c \quad (13)$$

The optimality condition for a Lagrangian approach in Eq. (9) consists of the equations

$$\begin{aligned} \mathbf{M} \ddot{\mathbf{x}} + \mathbf{C} \dot{\mathbf{x}} + \mathbf{K}(\mathbf{x} - \mathbf{X}) + \mathbf{G}^T \boldsymbol{\Lambda} &= \mathbf{F} \\ \mathbf{G} \mathbf{x} &= \mathbf{0} \end{aligned} \quad (14)$$

When the second condition in Eq. (8) is not satisfied for all contact elements, the contact constraint is no longer, and a collision-free finite element system must be solved. The finite element discretization of the variation of the Augmented Lagrange formulation in Eq. (9) yields

$$\begin{aligned} \mathbf{M} \ddot{\mathbf{x}}_{k+1} + \mathbf{C} \dot{\mathbf{x}}_{k+1} + \mathbf{K}(\mathbf{x}_{k+1} - \mathbf{X}) + \mathbf{G}^T \boldsymbol{\Lambda}_k \\ + \kappa(\mathbf{G}^T \mathbf{G} \mathbf{x}_{k+1}) = \mathbf{F}, \end{aligned} \quad (15)$$

where use is made of Eq. (14) and Eq. (6).

The temporal discretization scheme implemented for calculations is the Hilbert-Hughes-Taylor α -method. The parameters employed are $\alpha = -0.3$ and a time step increment $h = 50 \mu\text{s}$. These values give good accuracy and introduce advantageous

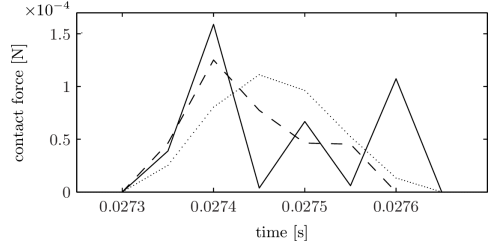


FIG. 2. Mediolateral coordinate of the contact force applied to an interface node. Solid line indicates the results for a Lagrange formulation; dotted line indicates the results for a penalty formulation with $\kappa = 10^7$; dashed line indicates an Augmented Lagrange formulation with 4 augmentations.

numerical damping. Further details can be found in [4].

III. RESULTS AND DISCUSSION

For all simulation, the tissue, geometry and initial conditions can be found in [4]. With regard to the augmented Lagrange technique, augmentations of a Lagrange multiplier associated to a slave node stop when the corresponding normal gap is less than 10^{-5} . The initial Lagrange multiplier vector is set to zero. The initial penalty parameter is 1, which increases by a factor of 10 when the total intersection volume is reduced by less than a 75% at each augmentation step.

The performance of different methods for contact constraint enforcement with regard to contact force estimations is illustrated in Fig. 2. The mediolateral component of the contact force applied to the interface node at initial position $(0.024, -0.136, -0.037)$ as a function of time is shown, for a subglottal pressure of 0.8 kPa. The results obtained by a penalty method with $\kappa = 10^7$ are shown in dotted line; the Lagrange multiplier method, in solid line; the Augmented Lagrange formulation with a maximum of 4 augmentations, in dashed line. Comparison between the penalty and Lagrange reaction force solution, makes apparent a spurious non-smooth behavior of the Lagrange multiplier solution. From physical considerations, a smooth transition at each time step may be expected. Consequently, the Lagrange approach may lead to wrong estimations of the instantaneous contact force. In an effort to improve this unsatisfactory behavior, the Augmented Lagrange approach seems to have a regularization effect with 4 augmentations per time step.

Fig. 3 shows the maximum mediolateral component of the total contact force calculated from the summation over all nodal contact forces. Circles

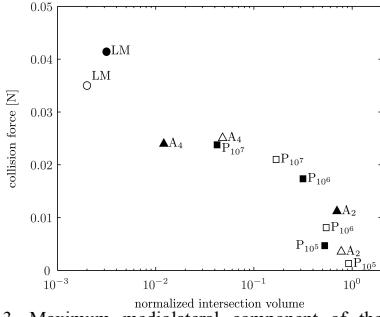


FIG. 3. Maximum mediolateral component of the total contact force as a function of the normalized intersection volume. The selected collision time interval is [0.036, 0.037], and the right vocal fold ($x > 0$) has been used for calculations.

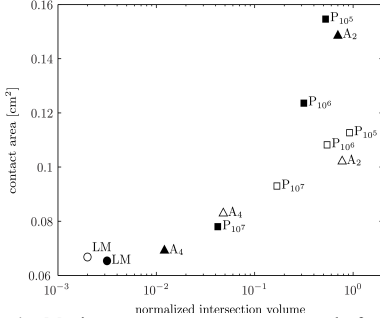


FIG. 4. Maximum contact area computed from the summation of interface triangles with any vertex loaded by a collision force. Symbols are the same as in Fig. 3.

indicate results for the Lagrange multiplier method (LM); squares correspond to a penalty formulation (P), where the subscript corresponds to the value of the penalty parameter; triangles correspond to the Augmented Lagrange approach (A), where the subscript indicates the maximum number of augmentations. Black and white marks show the results for a subglottal pressures of 0.8 kPa and 0.6 kPa, respectively. The horizontal axis corresponds to the interpenetration volume normalized to the maximum intersection volume when the effect of contact forces is neglected. The graph shows a clear effect of the violation of position-based contact constraint on contact force estimations. The Lagrange approach gives the smallest intersection volume, although, theoretically, the intersection volume should be zero. This small error is probably due to the contact detection algorithm. Furthermore, contact force computations with a penalty approach are highly dependent on the value of the penalty parameter. When

the subglottal pressure is modified, the Lagrange approach shows robustness in comparison with the penalty results. With regard to the Augmented Lagrange procedure, ideally the augmented multipliers are not dependent on the penalty parameter [3]. However, the numerical solution behaves differently, which may be due to the contact finite element computations. Robustness in the method may be introduced by enlarging the maximum number of augmentations. Nevertheless, exact contact force solution cannot be assured as the penalty parameter tends to infinity, and ill-conditioning of the system matrix may occur.

Fig. 4 shows the maximum contact area computed from the summation of interface triangles with any vertex loaded by a collision force. An influence of the interpenetration volume is apparent from the results. Again, the Lagrange multiplier method seems to be robust for subglottal pressure variations.

IV. CONCLUSIONS

Position-based contact constraints of vocal fold collision have been shown to have a clear effect on collision force and contact area estimations. The Lagrange multiplier method for contact constraint enforcement appears to be robust for pressure variations, but poor with regard to instantaneous contact force solution. An Augmented Lagrange approach with an Uzawa algorithm has a smoothing effect by introducing a penalty regularization term. However, the Penalty and the Augmented Lagrange results show strong dependency on penalty parameter choice. Alternative formulations of contact constraint may further improve contact force estimations.

REFERENCES

- [1] F. Alipour, D. A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," *J. Acoust. Soc. Am.*, vol. 108, pp. 3003–3012, 2000.
- [2] H. E. Gunter, "A mechanical model of vocal-fold collision with high spatial and temporal resolution," *J. Acoust. Soc. Am.*, vol. 113, pp. 994–1000, 2003.
- [3] P. Wriggers, *Computational contact mechanics*, John Wiley & Sons, 2002.
- [4] A. Granados, J. Brunskog, M. K. Misztal, V. Visseq, and K. Erleben "Finite element modeling of the vocal folds with deformable interface tracking," in *Proc. Forum Acust.* (Krakow, Poland), 2014.
- [5] N. Andreasson, and A. Evgrafov, and M. Patriksson, *Introduction to Continuous Optimization: Foundations and Fundamental Algorithms*, Studentlitteratur AB, 2006.

Paper C

Optical flow analysis of high-speed digital imaging of vocal fold vibration

Authors: Alba Granados, Jonas Brunskog

Submitted to: *IEEE Transactions On Biomedical Engineering.*

Optical Flow Analysis of High-Speed Digital Imaging of Vocal Fold Vibration

Alba Granados, Jonas Brunsog

Abstract—Feature extraction of high-speed videoendoscopy of the vocal fold vibrations provide valuable information for voice diagnosis. The present investigation presents a non-invasive procedure to analyze the two-dimensional vocal fold vibration captured by superior high-speed digital imaging. A dense optical flow algorithm is presented and adapted to the particular nature of videoendoscopic data. Motion discontinuities, temporal aliasing and image noise reduction are properly treated by robust statistics, regularization techniques and Gaussian filtering. Principal components analysis is performed to reduce data dimensionality. The algorithm is tested with synthetic and *in vivo* human high-speed videos. The procedure is shown to successfully capture the most significant superior modes of vibration and capable to quantify differences between normal and abnormal phonation.

Index Terms—Biomedical imaging, high-speed videoendoscopy, motion estimation, principal components analysis, vocal folds.

I. INTRODUCTION

Medical voice assessment is generally supported by diagnostic tools such as videoendoscopy of the vocal fold vibrations [1]. Objective evaluation of high-speed videoendoscopy, capable to record 2000–5000 frames per second, can be used to extract mechanical and dynamic properties of the vocal folds [2]–[4]. In order to perform a quantitative analysis of the images, different image processing procedures have been developed, namely glottal area segmentation algorithms [5], [6] and digital image correlation [3].

Extensive use has been made of segmentation algorithms, which has been accompanied by advances on image processing techniques [6], [7]. However, this procedure typically does not capture the two dimensional dynamics within the entire superior surface. In this sense, digital image correlation has arisen to provide a more extensive strain field, which can be used to extract detailed information of the stress field assuming a known tissue constitutive law [3], [8], and dense input feature may be of great advantage when dealing with continuum model inversion. Digital image correlation keeps track of a random pattern formed by black speckles that must be deposited on the tissue, which can be seen as a limiting factor when dealing with *in vivo* laryngoscopy data. Furthermore, spurious artifacts can be generated by an arbitrary choice of the speckle size [9], or by the use of naive algorithms that disregard the intricacies of the laryngoscopic images [5], such as spatial discontinuities or motion boundaries. In

contrast to image segmentation algorithms, state-of-the art digital image correlation algorithms for videoendoscopy presents weaknesses in some of this aspects.

An alternative non-invasive method closely related to image correlation is the optical flow. Based on image intensity tracking, optical flow algorithms can estimate the velocity field of moving images [10]. Recent investigations have applied optical flow algorithms to extract velocity vocal fold kymograms [11] or the superior two-dimensional motions [12]. In the present work, the formulation of the optical flow equations are adapted to the particular nature of vocal fold videoendoscopic data. For purposes of biomechanical parameters inference, a low error level of the extracted image data is desirable [4]. Videoendoscopy of the vocal fold present challenging scenarios such as occlusion of the intraglottal region, motion discontinuities due to the tissue wave-like dynamics and the vocal fold collision, and specular reflections of the camera light from the mucosal layer. Therefore, special effort is put to reduce the numerical error in the optical flow estimates by introducing robust estimators to preserve realistic features of the images, quasi-convex minimization, and anti-aliasing methods. Furthermore, the resulting optical flow is compressed using principal component analysis to preserve the most significant features in the vibration pattern.

The outline of the paper is as follows. In Section II the optical flow algorithm is presented, including numerical aspects and the principal components formulation used for analysis. Section III is devoted to the setup for the simulation results, followed by Section IV, which includes validation of the optical flow algorithm, results and discussion. Finally, in Section V the conclusions are presented.

II. METHODOLOGY

A. Optical flow estimation

Optical flow is the estimate of the projected two-dimensional velocity field of moving objects captured in an image sequence [10]. Let a pixel located at a position $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ have brightness $I(\mathbf{x}, t) \in \mathbb{Z}^+$ at time $t \in [0, N]$, and unknown optical flow $\mathbf{u} = (u_1, u_2) \in \mathbb{R}^2$. The time step between frames is denoted by δt . The main assumption of the optical flow theory is that brightness remains constant in time. That is,

$$I(\mathbf{x}, t) = I(\mathbf{x} + \mathbf{u}\delta t, t + \delta t). \quad (1)$$

for all pixels. A gradient-based continuity equation can be derived from the Taylor series expansion of the intensity

Submitted to IEEE Transactions on Biomedical Engineering
Dated: 13 May 2016.

A. Granados and J. Brunsog are at Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby DK-2800, Denmark. e-mails: albagranados@gmail.com and jbr@elektro.dtu.dk

function around (\mathbf{x}, t) [13] which, combined with Eq. (1), yields the data error equation

$$e_D(\mathbf{x}, \mathbf{u}) \equiv \nabla I(\mathbf{x}, t) \cdot \mathbf{u} + I_t(\mathbf{x}, t) = 0, \quad (2)$$

where the nabla operator is the spatial image gradient $\nabla I(\mathbf{x}, t) = (I_{x_1}(\mathbf{x}, t), I_{x_2}(\mathbf{x}, t))$, and the subscripts indicate partial derivatives with respect x_1 , x_2 and t . It is worth noticing that small velocity fields are assumed. Furthermore, the time dependent optical flow \mathbf{u} may be denoted by \mathbf{u}^t throughout the paper. Similarly, the intensity field may be denoted by I^t for simplicity.

The continuity equation in Eq. (2) is underdetermined, and a least-square estimate can be found by enforcing spatially constant velocity field \mathbf{u} within an image region $\{\mathbf{x}_s\}_{s \in \mathcal{R}}$ of pixels indexed by \mathcal{R} . This estimate can be calculated by the gradient expression to find the critical points. However, the naive solution to the problem may not be a good estimate if the system matrix is ill-conditioned. This occurs in regions with an approximately parallel intensity gradient field, or when a large amount of noise is present [14]. In the present case of high-speed videoendoscopy, images are typically noisy and have flatter image regions. To overcome this numerical difficulty, the size of \mathcal{R} can be increased. However, this may lead to large data error in Eq. (2). Alternatively, spatial coherence can be assumed, that is, assume a spatially smooth velocity field. In that case, the total optical flow $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_{n'})^T$ computed at each pixel position within a region $\mathcal{S} \supseteq \mathcal{R}$ of n' image pixels is the solution to the minimization problem

$$\min_{\mathbf{U}} E_D(\mathbf{U}) = \min_{\mathbf{U}} \sum_{s \in \mathcal{S}} \rho_D(e_D(\mathbf{x}_s, \mathbf{u}_s)) \quad (3a)$$

$$\text{s.t. } E_S(\mathbf{U}) = \sum_{s \in \mathcal{S}} (\rho_S(\nabla u_{s1}) + \rho_S(\nabla u_{s2})) \leq \varepsilon, \quad (3b)$$

with $\rho_D = \rho_S = \rho$ being the least-square error norm $\|\cdot\|_2^2$, ε an upper bound, and the nabla operator refers to the spatial gradient. The constrained minimization problem in Eq. (3) can be formulated as an unconstrained problem via Lagrangian formulation [15]

$$\min_{\mathbf{U}} E(\mathbf{U}) = \min_{\mathbf{U}} \{E_D(\mathbf{U}) + \lambda E_S(\mathbf{U})\}, \quad (4)$$

where $\lambda \geq 0$ is the regularization parameter that controls the importance given to both terms, and the minimizer is called the Tikhonov regularization solution. The regularization parameter must roughly approximate the measurement noise [16], and can be ideally determined by means of parameter-choice techniques.

1) *Robust estimation*: The spatial coherence assumption in Eq. (3b) is violated where flow discontinuity occurs, e.g., at motion and surface boundaries or when specular reflections are present. Similarly, the data error in Eq. (2) can be large at motion boundaries, when occlusion occurs or where significant of noise is present [17]. The least-square formulation of the minimization problem with a quadratic ρ -function gives each spatial point a weight proportional to its error value. This can lead to oversmoothed estimates of the optical flow when the movement of the object is described by a complex velocity field in certain regions of the image. Robust statistics provides

the framework to preserve realistic discontinuities by giving low weight to “outlier” positions with velocity singularities by using different ρ -functions. In this work, a fully robust approach is employed [17], where $\rho_D = \rho_S = \rho$ being the Lorentzian function

$$\rho(e, \tau) = \log \left(1 + \frac{1}{2} \frac{\|e\|_2^2}{\tau^2} \right), \quad (5)$$

where τ is a scale parameter that controls the convexity of the function. The derivative of ρ reflects the weight that is given to a particular position.

Nevertheless, the resulting minimization problem in Eq. (4) with ρ in Eq. (5) is non-convex. Hence, gradient-based algorithms may fail. To find a global minimum, a graduated non-convexity scheme (GNC) as in [18] can be applied. The minimization starts with a convex minimization problem where the argument is the quadratic approximation $E_{quad}(\mathbf{U})$ of $E(\mathbf{U})$ ($\rho = \|\cdot\|_2^2$) as

$$\min_{\mathbf{U}} \{\alpha E_{quad}(\mathbf{U}) + (1 - \alpha) E(\mathbf{U})\} \quad (6)$$

with $\alpha = 1$. The minimization is repeated with a lower value of $\alpha \in [1, 0]$, and outliers are iteratively introduced.

2) *Aliasing*: High amplitude velocity fields may cause problems with regards to the small velocity field assumption in Eq. (2) and temporal aliasing [10]. A coarse-to-fine strategy is followed, which estimates the velocity field in an iterative procedure that may be summarized as follows. The main process starts by sequentially blurring the image with a Gaussian filter of a certain support and standard deviation. Formally, a scale-space representation of an image signal is a one-parameter family of signals derived from the original signal as [19]

$$L(\mathbf{x}; \sigma^2) = \int_{\xi \in \mathbb{R}^2} I(\mathbf{x} - \xi) g(\xi; \sigma^2) d\xi = g(\mathbf{x}; \sigma^2) * I, \quad (7)$$

where

$$g(\mathbf{x}; \sigma^2) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|\mathbf{x}\|_2^2}{2\sigma^2}} \quad (8)$$

is the Gaussian kernel and σ^2 is the variance or scale parameter which regulates the amount of blurring. Because the family of kernels form a semi-group structure, a Gaussian pyramid of P_{levels} levels can be generated from the original grayscale image I by evaluating $L(\mathbf{x}; k\sigma^2)$ for $k = 1, \dots, P_{levels}$, for a given standard deviation.

At coarser scales, the magnitude of the estimated velocities decreases due to the blurring. Then, the coarse-to-fine strategy continues by calculating the optical flow in the coarsest scale from an initial optical flow estimate. Assuming constant brightness to the resulting image, a residual flow can then be calculated as the solution to the minimization problem in Eq. (6). The linearized gradient numerical approximations to solve the problem can be found in [20]. With the resulting residual flow, the optical flow can be updated and used as the initial optical flow in the next blurred image in the Gaussian pyramid with a lower standard deviation, and the process is repeated.

The overall procedure to estimate the optical flow between two consecutive frames can be summarized as in Algorithm 1.

Algorithm 1 Robust optical flow estimation

Input: image I^t ; image $I^{t+\delta t}$; ρ_D ; ρ_S ; λ ; τ_D ; τ_S ; σ ; P_{levels} ;
 GNC_{steps} ;
Output: $\mathbf{U}^t = (\mathbf{u}_1^t, \dots, \mathbf{u}_{n'}^t)^T$
1: convert to grayscale I^t and $I^{t+\delta t}$
2: create Gaussian filters with σ and support $[-(\sigma+0.5), (\sigma+0.5)] \times [-(\sigma+0.5), (\sigma+0.5)]$.
3: construct image pyramid, filter recursively and rescale at ration $1/(2\sigma^2)$ [21]
4: set initial optical flow $\mathbf{U}^t = 0$.
// correct non-convexity of Eq. (4) for robust estimation:
5: **for** each GNC step **do**
// correct possible aliasing:
6: **for** each pyramid level from coarse to fine **do**
7: set residual flow $\delta \mathbf{U}^t = 0$
8: compute spatial and temporal partial derivatives based on Eq. (6) [20]
9: solve for $\delta \mathbf{U}^t$
10: update flow fields $\mathbf{U}^t = \mathbf{U}^t + \delta \mathbf{U}^t$
11: **end for**
12: decrease α linearly to solve Eq. (6) in the next iteration
13: **end for**
14: **return** \mathbf{U}^t

B. Principal components analysis

Principal component analysis (PCA) allows a simpler understanding and visualization of multivariate data by reducing the dimensionality of the variables. The result is a change of basis or principal components that maximizes the variance of the data. For an endoscopic image sequence with length N , PCA can be performed by interpreting the total optical flow estimate \mathbf{U} as a multivariate random variable, and N the number of realizations. The optical flow at each time step is wished to be represented as a linear combination of a number of flow basis [22]. This linear combination is determined by a number of coefficients corresponding to the dimensionality reductions, which summarize most of the information that this particular instant optical flow contains. In this sense, redundant pixel-wise optical flows are eliminated.

Let the optical flow between two consecutive frames at times t and $t+\delta t$ be represented as an $n = 2n' = 2WH$ vector $\mathbf{U}^t = (u_{11}, \dots, u_{n'1}, u_{12}, \dots, u_{n'2})^T$, with W and H being the number of pixels in the horizontal and vertical directions, respectively. Then, the optical flow between two frames can be expressed as a summation of r principal components $\mathbf{U}_{pc} \in \mathbb{R}^n$ as

$$\mathbf{U}^t \approx \sum_{k=1}^r \alpha_k^t \mathbf{U}_{pc}^k \in \mathbb{R}^n \quad (9)$$

with $r < n$. Hence, the optical flow $\mathbf{U}^t \in \mathbb{R}^n$ may be represented by a reduced optical flow

$$\tilde{\mathbf{U}}^t = (\alpha_1^t, \dots, \alpha_r^t)^T \in \mathbb{R}^r \quad (10)$$

Principal component decomposition seeks the linear transformation

$$\tilde{\mathbf{U}}^t = \mathbf{M} \mathbf{U}^t \quad (11)$$

with $\mathbf{M} \in \mathbb{R}^{r \times n}$ that maximizes the variance of N realizations of the n -dimensional random variable \mathbf{U} which is here assumed to be zero-mean. As $r < n$, classical eigenvalue factorization can not be used, and singular value decomposition (SVD) may be performed instead [23]. It can be seen that, if $\mathbf{U} \in \mathbb{R}^{N \times n}$, $\mathbf{V} \in \mathbb{R}^{n \times N}$ and the diagonal matrix $\mathbf{G} \in \mathbb{R}^{N \times N}$ with decreasing elements or singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N$ are the singular value decomposition of an $N \times n$ -dimensional matrix containing all the data with $N < n$ and $r = N$, then

$$\mathbf{M}^T = \mathbf{V}$$

maximizes the variance and the column vectors of \mathbf{V} are the principal components \mathbf{U}_{pc} . The variances correspond to the squared singular values and the first k principal components describe the

$$\frac{\sigma_1^2 + \dots + \sigma_k^2}{\sigma_1^2 + \dots + \sigma_k^2 + \dots + \sigma_N^2} \cdot 100 \quad (12)$$

percent of the total variance of the original optical flow. Note that if $r < n$, the system in Eq. (11) is underdetermined, and the solution is the (naive) least-square solution. Finally, the endoscopic image sequence would generally not be zero-mean. By defining the mean optical flow of an N image sequence as

$$\bar{\mathbf{U}} = \frac{1}{N} \sum_{t=1}^N \mathbf{U}^t \quad (13)$$

It is easy to see that

$$\tilde{\mathbf{U}}^t = \mathbf{M} (\mathbf{U}^t - \bar{\mathbf{U}}), \quad (14)$$

and the transformation matrix reminds the same.

In summary, once the optical flow is estimated following the procedure described in the previous section, a matrix $N \times 2WH$ dimensional is constructed. Because typically $N \ll WH$, then singular value decomposition is performed. Hence, the time-varying optical flow of an endoscopic image sequence can be represented by a time-varying r -dimensional vector as in Eq. (9). For small r , simplicity is obtained in the principal component representation.

III. SIMULATION SETUP

The performance of the optical flow algorithm is investigated with high-speed videos of *in vivo* human vocal folds as well as synthetic replicas for validation purposes. The recording of the vocal fold replica corresponds to the EPI model in [24]. The model consists of a silicone three-layer bodycover model similar to [25] covered by a highly flexible thin epithelium layer which promotes inferior-superior mucosal wave-like motion and a convergent-divergent profile. The center of the ligament layer contains a tensioned isotropic acrylic fiber which restricts the inferior-superior vibratory motion. The image sequence contains 30 frames and captures steady vocal fold oscillations at a phonation onset of $1.2 \cdot 0.24$ kPa and vibration frequency about 100 Hz, corresponding to typical male speech values. A digital camera with a frame rate of 3000 frames per second and a resolution of 720×960 pixels is used to capture the grayscale image sequence, although

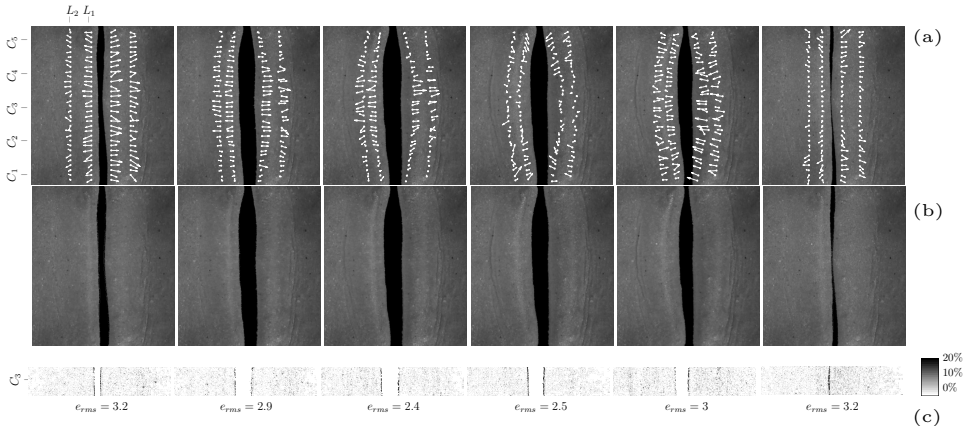


Fig. 1. From left to right, image sequence of the synthetic vocal fold model at $t = 1, 6, 9, 12, 15$ and 28 . (a) Estimated optical flow of the image $I(\mathbf{x}, t)$ at a selected number of pixels. The optical flow magnitudes has been rescaled for clarity reasons. (b) Warped image $I'(\mathbf{x}, t + \delta t)$ as in Eq. (15). (c) Pixel percent error e_{pix} as defined in Eq. (17) within the range $0 - 20\%$ (grayscale colorbar at the bottom right). The total e_{rms} is also shown.

the optical flow algorithm is applied to a manually selected reduced region of interest.

Digital high-speed imaging of human vocal fold steady vibrations is performed from one voice-healthy subject (male, age 54 ('HN')), one subject with a functional voice problem (male, unknown age with mucus before throatcleaning ('F')) and one subject with an organic voice problem (female, age 42 with vocal fold paresis or weakness ('O')). The subjects undertook sustained phonation in normal pitch. Additionally, the male healthy subject phonated at higher sound pressure level and fundamental frequency (maximum 94 dB sound pressure level at 255 Hz ('HL')). The image sequence is captured with a digital camera (HRES Endocam, model 5562.9 color; Wolf, Germany) at 2000 frames per second and contains about 300 frames. The image resolution is 256×256 pixels, but as with the synthetic data, the optical flow algorithm has been applied to a reduced image area.

A MATLAB routine is used to convert (R,G,B) banded color images to a gray-scale intensity according to the transformation $0.2989R + 0.5870G + 0.1140B$. Parts of the code registered in the Copyright 2007-2010, Brown University, Providence, RI, USA are used to estimate the optical flow. Based on [21], the graduated non-convexity optimization scheme stops after $GNC_{steps} = 3$ iterations, and $\tau_S = 0.1$ and $\tau_D = 3.5$ correspond to the selected scale parameter for the robust function of the spatial and data terms, respectively. The Gaussian pyramid is generated from multiples of the variance $\sigma^2 = 0.16$. Unless otherwise stated, the robust functions for both the data and spatial terms is the Lorentzian function in Eq. (5), the number of Gaussian pyramids in the anti-aliasing process is $P_{levels} = 4$, and the regularization parameter is chosen to be $\lambda = 0.5$.

IV. RESULTS AND DISCUSSION

A. Method evaluation

The performance of the method is evaluated with high-speed videos of the synthetic vocal fold model, thereby reducing possible effects of camera movements. For simplicity, it is here assumed $\delta t = 1$, and consequently t refers to the frame number. Fig. 1(a) shows the results during an oscillation cycle at frames $t = 1, 6, 9, 12, 15$ and 28 , from left to right. The computed optical flow is illustrated at a reduced number of pixel, for clarity reasons. Visual evaluation of the optical flow estimates reveals realistic deformations: the mid-coronal plane (around the line C_3) undertakes large displacements, showing large magnitudes during the opening phase ($t = 6$ and 9), and small magnitudes at maximum opening and closing ($t = 12$ and 28), in agreement with previous theoretical results [26]. This is illustrated in Fig. 2 for a number of pixels at five anterior-posterior positions, C_1 to C_5 , along the lines L_1 and L_2 (see Fig. 1(a) left). The figure shows the maximum medio-lateral displacement computed from the pixel trajectory estimates. The results reveal small displacements in the anterior and posteriormost lines (C_1 and C_5). Furthermore, as the

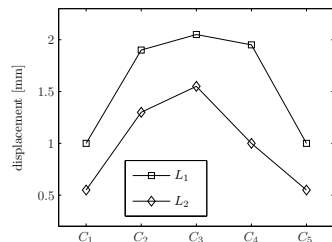


Fig. 2. Maximum lateral displacement of the pixel trajectories calculated from the optical flow estimation for pixels initially located at the coordinates shown in Fig. 1(a) left.

pixel position approaches the glottal midline, the maximum displacement increases, in agreement with the results reported in the literature [27]. Finally, the maximum displacements is 2.05 mm, slightly lower than the value 2.27 mm reported in [24] for the same synthetic model. This may be explained by the choice of the line L_1 arbitrarily close to the midline or by uncertainties in the pixel size calculation.

In order to assess the accuracy of the optical field estimates, the root-mean-square error e_{rms} between a warped image and the actual images is calculated [28]. Given a velocity field estimate \mathbf{u} at a pixel at the position \mathbf{x} , the warped image is defined as

$$I'(\mathbf{x}, t + \delta t) = I(\mathbf{x} - \mathbf{u}\delta t, t). \quad (15)$$

Then, the root-mean-square error at each time instant t is

$$e_{rms} = \sqrt{\frac{\sum_{s \in S \subset \mathcal{I}} (I(\mathbf{x}^s, t) - I'(\mathbf{x}^s, t))^2}{|S|}}, \quad (16)$$

where a portion $S \in \mathcal{I}$ of the original image area \mathcal{I} is considered to avoid evaluation of the intensity field outside the domain. In this work, the region S has been selected to represent the centered 94% of the original area through a trial-and-error process. To visualize the pixel-wise optical flow performance, the percent error between the actual frame at $t + \delta t$ and the warp image in Eq. (15) is calculated as

$$e_{pix} = \frac{I(\mathbf{x}, t + \delta t) - I'(\mathbf{x}, t + \delta t)}{I(\mathbf{x}, t + \delta t)} \cdot 100 \quad (17)$$

at each pixel position with $I(\mathbf{x}, t + \delta t) \neq 0$. Fig. 1(b) shows the warped image $I'(\mathbf{x}, t + \delta t)$ and Fig. 1(c) the corresponding root-mean-square error e_{rms} in an anterior-posterior area that contains C_3 . As stated earlier, high-speed movies of the vocal fold vibrations capture motion discontinuities at collision and at the vocal fold edge, and due to the mucosal wave-like behavior. From Fig. 1(c), it is clear that large pixel-wise errors e_{pix} occur at the vocal fold edge. Note that at maximum opening ($t = 12$) the error e_{rms} decreases, which may be explained by the small magnitudes of the velocity field estimates at this phase. At vocal fold collision ($t = 28$) the algorithm fails at finding good optical flow estimates, illustrated by the large values of the errors e_{rms} and e_{pix} . In addition, notice that at the pixel-wise error in Fig. 1(c) at during vocal fold opening ($t = 9$) and closing ($t = 15$) is slightly larger within the solid domain, compared to the error computed at other phases of the glottal cycle. This may be possibly due to the wave-like motion.

TABLE I
ROOT-MEAN-SQUARE ERROR e_{rms} OF THE SYNTHETIC MODEL.

λ	ρ	P_{levels}	mean(e_{rms})	std(e_{rms})
$\lambda = 0.01$	L	2	3.1085	0.3682
$\lambda = 0.1$	L	2	3.0340	0.3494
$\lambda = 0.5$	L	2	2.9644	0.3218
$\lambda = 0.5$	L	4	2.804	0.3105
$\lambda = 0.5$	L	6	2.7985	0.3088
$\lambda = 0.5$	Q	4	2.816	0.3111

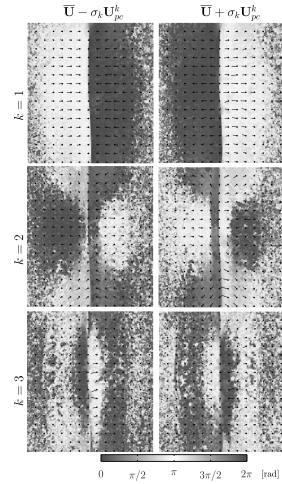


Fig. 3. For the synthetic vocal fold model, illustration of the space generated by the first three principal components \mathbf{U}_{pc}^1 , \mathbf{U}_{pc}^2 and \mathbf{U}_{pc}^3 (rows) within the domain $[-\sigma_k, \sigma_k]$ (left and right columns, respectively).

In order to evaluate the performance of the robust estimation of the optical flow, the root-mean-square error is computed with a quadratic norm ($\rho = \|\cdot\|_2^2$, denoted as ‘Q’), and the Lorentzian function in Eq. (5) (denoted as ‘L’). In Table I the mean and standard deviation are calculated over 30 time consecutive frames. Notice the positive effect of robust estimation against classical least-square estimates. This may indicate the presence of complex motion and demonstrate the need of proper error analysis in intensity-based optical flow algorithms applied to high-speed videos of the vocal fold vibration. Furthermore, the effect of temporal aliasing is evaluated by modifying the number of levels of the Gaussian pyramid, and the effect of smooth displacement enforcement by setting different values of the regularization parameter λ . The results are shown in the same table, and reveal a decrease of e_{rms} as more regularization is introduced (increasing λ). This indicates that the presence of image noise is significant. However, very large λ may lead to over-smoothed optical flow estimations, and realistic motion discontinuities, such as the wave-like behavior, may not be captured by the algorithm. Thus, an arbitrary upper bound is set to $\lambda = 0.5$. Parameter-choice techniques may be applied to automatically select the optimal regularization parameter, although this is left for possible future investigations. Furthermore, the effect of temporal aliasing is clear from the results shown in the table. Gaussian filters seem to be able to reduce the amplitudes of the estimated vocal fold velocity field and thereby reduce the root-mean-square error. Finally, although not shown here for the sake of brevity, peaks of e_{rms} occur at collision, as Fig. 1(c) ($t = 1$ and 28) indicates. Therefore, low standard deviations may indicate a decrease of optical flow estimation error at collision.

value of the coefficient α_2 , the flow phase at the anterior and posterior extremes tends to push the vocal fold together with a strong anterior-posterior vector component, while the flow within the mid region (C_3) pulls the vocal folds apart (see Fig. 3 for an enlarged image and note the light gray color). Finally, during vocal fold closing, the second principal component does not seem to have a significant effect on the motion, while the third component plays a role. Great part of the total motion occurs close to the midline. Due to the change of sign of α_3 , the mode appears to pull apart when the closing phase starts, and before collision the mode pushes them together. For positive values of α_1 (push together), this observation may reveal a transition from negative strains to positive strains during closure, which is in agreement with the results from image correlation in [3]. Finally, note a slight asymmetry with regard to optical flow magnitude in the third component, suggesting asymmetries at collision (see Fig. 1(a) last frame at $t = 28$).

With regard to human high-speed videos of the vocal fold vibration, the principal component coefficients in Fig. 4(b) of the voice-healthy data 'HN' show a similar trend compared to the results of the synthetic model. 82% of the variance in the image sequence is accounted by the first three principal components. Notice that the graph appears to be mirrored with respect the zero value. However, the posterior extreme of the first optical flow shows how the vocal folds are pulled apart during opening. Therefore, the previous discussion reminds the same. Furthermore, the strong anterior-posterior motion is captured by the anterior extreme of the same component (light gray). Nevertheless, the vibration pattern decomposition is different, as the second mode seems to be significant at collision, and the third component at maximum opening. Furthermore, the incomplete closure (see image frame at the bottom right of the figure) appears to be captured by the second component. Finally, notice that the principal modes of vibration are captured despite the presence of camera reflexions on the mucosal superficial layer.

2) *Abnormal phonation*: The high-speed videos of vocal fold vibration of subjects with functional ('F') and organic ('O') voice problems, and healthy voice at loud and high pitch ('HL') are analyzed. Fig. 5 shows the time-varying principal component coefficients for the three subjects. The general trend is different from the results from normal phonation. The curves in Fig. 5(a) correspond to the same voice-healthy subject in Fig. 4(b). The 82% of the variation is explained by six principal components, which indicates an increase of the complexity of the vibration pattern in loud and high-pitch voices. A second component has a strong effect on the main vibration pattern, in contrast to previous results of normal phonation where only one mode explains most of the motion variance. From video observations, an anterior-posterior temporal delay at maximum opening is apparent, which may be related to the strong presence of the second component when the vocal fold closure starts. Note that the general trend of the third and forth components is very similar to the second and third components for normal phonation. At contact, the third component plays a similar role to the second component in normal phonation. Finally, the mode of

phonation at loud and high-pitch voice of the specific subject results into the formation of a posterior chink previous to collision, captured by the third component (top right in the figure).

In Fig. 5(b) the results of the subject with vocal fold mucus before throatcleaning ('F') is shown. These components explain 69% of the total variance. The presence of sticky mucus seems to cause a non-uniform vibration pattern, demonstrated by the two different significant modes during opening (second component) and closing (first component). The third component active at contact also shows asymmetries. Furthermore, after throatcleaning the coefficients curve was very similar to the curves in Fig. 4, although not shown here for the sake of brevity.

Fig. 5(c) corresponds to the results of the subject with right vocal fold paresis (or weakness). Even though the first three principal components only explain 31% of the total variance, probably due to the difficulties of the subject to sustain vocal steady vibrations, these components are enough to illustrate the clear asymmetry during vibration. Two modes explain the left vocal fold motion, and one mode explains the motion of the anterior extreme of the weak vocal fold, in agreement with video observations. The incomplete closure of the vocal fold is in agreement with the optical flow results.

V. CONCLUSIONS

An optical flow algorithm has been successfully applied to high-speed movies of the vocal fold vibration to track vocal fold deformation as an alternative to existing invasive techniques. The complex motion capture by the image sequence has been shown to benefit from robust statistics, regularization and anti-aliasing techniques. Nevertheless, the algorithm has shown weaknesses at deformation at collision. Principal component analysis of the dense optical flow estimates has been shown to preserve the most significant superior modes of vibration. A procedure consisting of the projection of the optical flows onto the principal component space has been shown to successfully quantify differences between normal and abnormal phonation.

ACKNOWLEDGMENT

This work was supported by grant number 110230 from the Swedish organization AFA Försäkring. The authors would like to express their gratitude to Dr. Scott Thomson (Brigham Young University, Provo, United States) for sharing the recordings of the synthetic vocal fold models, and Dr. Roland Rydell (Lund University, Lund, Sweden) for providing the high-speed videos of the vocal fold vibrations.

REFERENCES

- [1] M. Hirano and D. M. Bless, *Videostroboscopic Examination of the Larynx*. Singular Publishing Group, 1993.
- [2] C. Tao, Y. Zhang, and J. J. Jiang, "Extracting physiologically relevant parameters of vocal folds from high-speed video image series." *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 794–801, may 2007.
- [3] M. Spencer, T. Siegmund, and L. Mongeau, "Determination of superior surface strains and stresses, and vocal fold contact pressure in a synthetic larynx model using digital image correlation." *J. Acoust. Soc. Am.*, vol. 123, no. 2, pp. 1089–1103, feb 2008.

- [4] A. P. Pinheiro, D. E. Stewart, C. D. MacIel, J. C. Pereira, and S. Oliveira, "Analysis of nonlinear dynamics of vocal folds using high-speed video observation and biomechanical modeling," *Digital Signal Process.*, vol. 22, no. 2, pp. 304–313, 2012.
- [5] T. Wittenberg, M. Moser, M. Tigges, and U. Eysholdt, "Recording, Processing, and Analysis of Digital High-speed Sequence in Glottography," *Mach. Vision Appl.*, vol. 8, pp. 399–404, 1995.
- [6] J. Lohscheller, H. Toy, F. Rosanowski, U. Eysholdt, and M. Döllinger, "Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos," *Med. Image Anal.*, vol. 11, no. 4, pp. 400–413, 2007.
- [7] C. Manfredi, L. Bocchi, G. Cantarella, and G. Peretti, "Videokymographic image processing: Objective parameters and user-friendly interface," *Biomed. Signal Proces.*, vol. 7, no. 2, pp. 192–201, 2012.
- [8] M. Sutton, C. Mingqi, W. Peters, Y. Chao, and S. McNeill, "Application of an optimized digital correlation method to planar deformation analysis," pp. 143–150, 1986.
- [9] P.-C. Hung and A. S. Voloshin, "In-plane strain measurement by digital image correlation," *J. Braz. Soc. Mech. Sci. Eng.*, vol. 25, no. 3, pp. 215–221, 2003.
- [10] D. Fleet and Y. Weiss, "Optical Flow Estimation," in *Handbook of Mathematical Models in Computer Vision*, N. Paragios, Y. Chen, and O. Faugeras, Eds. Springer US, 2006, ch. 15, pp. 237–257.
- [11] G. Andrade-Miranda, N. Henrich Bernardoni, and J. I. Godino-Llorente, "Optical-flow kymograms and glottovibrograms : a new way to present high-speed data for laryngeal assessment," in *9th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Application*, Firenze, Italy, 2015, pp. 71–74.
- [12] A. Granados and J. Brunsog, "Inverse problem in high-speed recordings of the vocal folds," in *Book of Abstracts of the 11th Pan-European Voice Conference*, Florence, Italy, 2015.
- [13] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Imaging*, vol. 130, pp. 674–679, 1981.
- [14] J. K. Kearney, W. B. Thompson, and D. L. Boley, "Optical flow estimation: an error analysis of gradient-based methods with local optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 2, pp. 229–244, 1987.
- [15] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1-3, pp. 185–203, 1981.
- [16] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*. Springer-Verlag New York, 2005.
- [17] M. J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," *Comput. Vis. Image Und.*, vol. 63, no. 1, pp. 75–104, 1996.
- [18] D. Sun, S. Roth, J. P. Lewis, and M. J. Black, "Learning optical flow," in *European Conference on Computer Vision*, 2008, pp. 83–97.
- [19] T. Lindeberg, "Scale-space: A framework for handling image structures at multiple scales," in *Proc. CERN School of Computing*, Egmond aan Zee, The Netherlands, 1996.
- [20] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. 8th European Conference on Computer Vision*. Springer, 2004, pp. 25–36.
- [21] D. Sun, S. Roth, and M. J. Black, "A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles Behind Them," *Int. J. Comput. Vision*, vol. 106, no. 2, pp. 115–137, 2013.
- [22] D. J. Fleet, M. J. Black, Y. Yacoob, and A. D. Jepson, "Design and use of linear models for image motion analysis," *Int. J. Comput. Vision*, vol. 36, no. 3, pp. 171–193, 2000.
- [23] R. E. Madsen, L. K. Hansen, and O. Winther, "Singular value decomposition and principal component analysis," *Neural Netw.*, vol. 1, pp. 1–5, 2004.
- [24] P. R. Murray and S. L. Thomson, "Vibratory responses of synthetic, self-oscillating vocal fold models," *J. Acoust. Soc. Am.*, vol. 132, no. September, pp. 3428–3438, 2012.
- [25] I. R. Titze, S. S. Schmidt, and M. R. Titze, "Phonation threshold pressure in a physical model of the vocal fold mucosa," *J. Acoust. Soc. Am.*, vol. 97, no. 5 Pt 1, pp. 3080–3084, 1995.
- [26] M. D. O. Rosa, J. C. Pereira, M. Grellet, and A. Alwan, "A contribution to simulating a three-dimensional larynx model using the finite element method," *J. Acoust. Soc. Am.*, vol. 114, no. 5, pp. 2893–2905, 2003.
- [27] A. Boessenecker, D. A. Berry, J. Lohscheller, U. Eysholdt, and M. Doellinger, "Mucosal wave properties of a human vocal fold," *Acta Acust. united Ac.*, vol. 93, no. 5, pp. 815–823, 2007.
- [28] T. Lin and J. Barron, "Image reconstruction error for optical flow," in *Vision Interface (VI1995)*, 1994, pp. 73–80.
- [29] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice*, ser. Speech Science. Singular Thomson Learning, 2000.
- [30] F. Alipour, D. A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," *J. Acoust. Soc. Am.*, vol. 108, no. 6, pp. 3003–3012, 2000.

Paper D

**Image-based parameter
identification of a finite
element model of the vocal
folds**

Authors: Alba Granados, Jonas Brunskog

Unfinished manuscript

Image-based parameter identification of a finite element model of the vocal folds[☆]

Alba Granados^{a,*}, Jonas Brunskog^a

^a*Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby DK-2800, Denmark*

Abstract

High-speed movies of the vocal fold vibration are valuable data to infer vocal fold characteristic for clinical purposes, where uncertainty information of the estimates can contain important information. This work presents the theoretical study of a non-invasive Bayesian method to determine vocal fold mechanical features of a self-sustained three-dimensional continuum model of the vocal folds, as well as a thorough discussion. A linear and Gaussian state-space model is derived, which accounts for numerical and measurement errors. The evolution model is based on a constrained finite element model of the vocal folds, and the observation model on a dense optical flow algorithm. Kalman filtering is used to formulate a maximum likelihood algorithm for single-valued parameter estimation based on the expectation-maximization algorithm. It is shown that the problem faces a number of computational challenges which require further investigations in order to evaluate the proposed statistical method.

Keywords: Vocal folds, stochastic finite element modeling, optical flow, bayesian state-space identification, kalman filtering, expectation-maximization algorithm, mesh generation.

1. Introduction

Vocal fold tissue parameters have an influence on the mechanics of the vocal folds. Given a mathematical model of the vocal folds, relevant features of the system can be investigated by enforcing changes in the tissue parameters (forward problem) (Steinecke and Herzel, 1995; Zhang and Jiang, 2008), and by inferring tissue parameters from measurement data (inverse problem) (Yang et al., 2011; Pinheiro et al., 2012). The present work focuses on the latter approach based on laryngeal endoscopic clinical observations. The mathematical formulation presented here aims at capturing the actual complexity of the problem with regard to biomechanical modeling, observation data extraction and inversion approach.

One-layered classical two-mass vocal fold models (Döllinger et al., 2002; Pinheiro et al., 2012; Cataldo et al., 2013) or multimass vocal fold models (Schwarz et al., 2008; Yang et al., 2011) are typically used in inversion procedures due to its simplicity and low dimensionality. However, these systems are deficient at modeling the actual layered structure of the vocal folds, which is crucial to explain the complex vocal fold motion (Spencer et al., 2008; Murray and Thomson, 2012). With the aim to extract more complex tissue properties, a simple two-layered structure can be introduced by body-cover models (Hadwin et al., 2016). However, as with the two-mass systems, the model does not capture the anterior-posterior geometry, which is necessary to explain the superior vocal fold observations. In an effort to model a realistic vocal fold tissue, in (Schmidt et al., 2011) a static linear elastic finite element model of the vocal folds is used. However, the proposed method can only be applied to *ex vivo* larynges, as it requires the measurement

[☆]Preliminary theoretical study. Not ready for submission.

*Corresponding author at: Technical University of Denmark, Ørstedes Plads, Building 352, Kgs. Lyngby DK-2800, Denmark. Tel.: +45 45 25 39 33.

Email addresses: albagranados@gmail.com (Alba Granados), jbr@elektro.dtu.dk (Jonas Brunskog)

of tissue deformation from applied loads. Hence, the observation data used in the method presents a big limitation for the extension of the method to *in vivo* clinical applications (Döllinger et al., 2002).

The laryngoscopic data is typically based on non-invasive techniques such as one-dimensional glottal area estimations (Pinheiro et al., 2012; Hadwin et al., 2016), time-varying contours of the segmented glottis (Schwarz et al., 2008) or one-dimensional medial lateral displacements (Döllinger et al., 2002). However, this data captures the motion only on the free vocal fold superior edge, which may not be sufficient to explain the inner mechanical properties of the vocal fold (Hadwin et al., 2016). However, recent studies have shown that detailed two-dimensional vocal fold superior dynamics can be captured during human phonation with optical flow algorithm (Andrade-Miranda et al., 2015; Granados and Brunskog, 2015). This opens the possibility to overcome the ill-posedness of inferring inner tissue parameters from low-dimensional input data.

Furthermore, the ill-posedness of the problem must cause poor estimates of the tissue parameters estimates if the inversion approach is not applied in a thoughtful way. As pointed out in (Schwarz et al., 2008), noisy data is present in image processing of vocal fold endoscopy, which leads to uncertainty affecting the results of the inversion. Deterministic least-squares methods are typically applied to resolve the inverse problem, which do not take into account measurement noise unless proper regularization techniques are applied. Furthermore, large modeling errors can occur due to the model simplifications and discretizations, which may lead to misleading solutions (Kaipio and Somersalo, 2005; Hadwin et al., 2016). In addition, vocal fold models are generally time dependent and physical quantities depend on previous states of the system. Finally, for clinical applications an uncertainty analysis of the results may be of importance before making decisions. Because direct methods have been shown to be partly unsatisfactory, the present work exploits the underlying statistical model for vocal fold parameter estimation by formulating a nonstationary state-space identification problem within a Bayesian inference framework (Kaipio and Somersalo, 2005). In an effort to assess the complexity of the vocal fold system, a continuum model is employed, as well as dense data from superior optical flow estimates.

This paper presents a theoretical framework to solve the continuum vocal fold inversion problem. A thorough discussion on the problem formulation and its limitations is presented. Throughout the paper, the computational challenges to be faced in the next step of the study are emphasized. Hence, no simulation results are yet available.

The outline of the paper is as follows. In Section 2 a brief review of nonstationary Bayesian statistical inversion theory and the main assumptions on which the vocal fold inference problem is based are presented. The state evolution and observation models are described and discussed, as well as the parameter inference method and the mesh generation procedure. Section 3 is devoted to the setup for future simulations. The potential of the method is illustrated and the necessary future work is discussed. Finally, in Section 4 the conclusions are presented.

2. Methodology

2.1. Nonstationary Bayesian statistical inverse problem

Statistical inference provides a general mathematical framework to extract features of a partially unknown system given the observed output of this system, taking into account uncertainties. In contrast to frequentist inference, Bayesian inference theory is built under the assumption that it exists prior knowledge on the system, and provides a solution to the system in the form of a posterior density distribution conditioned to the observations. In the case of vocal fold features estimation from visual inspection, estimates of the time evolution vocal fold oscillations are available via biomechanical models. These time-dependent models can be used as prior knowledge. Furthermore the observed image sequence is also dependent on time. Hence, the inverse problem can be formulated as a so-called state-space model typically used in time series analysis (Hamilton, 1994; Kaipio and Somersalo, 2005). In this work, vocal fold states are estimated at each time step conditioned to the knowledge of image observations up to that time step, and hence, the problem is a filtering problem. The mathematical formulation and main assumptions of the Bayesian filtering method for vocal fold features estimation is described in this section. Throughout this paper, random variables will

be denoted by non-bold uppercase letters, and the realization by lowercase letters. In addition, random variables are assumed to be absolutely continuous and normally distributed. The system is assumed to be linear.

Let $\{Z_k\}_{k \geq 0}$ be an unknown stochastic process (state vectors) at different time instants, and $\{Y_k\}_{k \geq 1}$ an observable stochastic process (observation vectors). A linear state-space model may be expressed as

$$Z_{k+1} = \mathbf{G}(\boldsymbol{\theta})Z_k + F_k + W, \quad W \sim \mathcal{N}(0, \Gamma_w) \quad (1a)$$

$$Y_k = \mathbf{H}Z_k + V, \quad V \sim \mathcal{N}(0, \Gamma_v) \quad (1b)$$

consisting of a state evolution model (1a) and an observation model (1b), with additive zero mean time-independent Gaussian noises denoted as W and V , respectively, mutually independent and also mutually independent of the initial state vector Z_0 . The vector $\boldsymbol{\theta}$ is an unknown parameter vector that affects the system behavior. In this work, the state process is assumed to be a first-order Markov process with known initial state Gaussian probability density $\pi(z_0)$. The observable process is assumed to be a first-order Markov process with respect to the state process. In addition, the Markov transition kernels $\pi(z_{k+1}|z_k)$ satisfy

$$\pi(z_{k+1}|z_k, y_1, \dots, y_k) = \pi(z_{k+1}|z_k), \quad (2)$$

and are determined by the evolution model in Eq. (1a). The likelihood function $\pi(y_k|z_k)$ is given by the observation model in Eq. (1b). Given a set of observations $y_{1:N} = \{y_1, y_2, \dots, y_N\}$, a filtering problem consists of finding an estimate of the posterior density $\pi(z_{k+1}|y_{1:k+1})$, that is, the conditional probability of the unknown state vector z_{k+1} conditioned on a number of measurements. A two-steps recursive scheme based on Bayes' formula is used to find the posterior density. At first, a time-evolution updating (prior) computes the prediction $\pi(z_{k+1}|y_{1:k})$ based on the transition kernels $\pi(z_{k+1}|z_k)$ which are dependent on the prior knowledge of the system according to the marginal density

$$\pi(z_{k+1}|y_{1:k}) = \int \pi(z_{k+1}|z_k) \pi(z_k|y_{1:k}) dz_k. \quad (3)$$

Then, an observation updating (posterior) estimates the posterior filtering distribution $\pi(z_{k+1}|y_{1:k+1})$ given the likelihood $\pi(y_{k+1}|z_{k+1})$ as

$$\pi(z_{k+1}|y_{1:k+1}) = \frac{\pi(y_{k+1}|z_{k+1}) \pi(z_{k+1}|y_{1:k})}{\pi(y_{k+1}|y_{1:k})}, \quad (4)$$

where

$$\pi(y_{k+1}|y_{1:k}) = \int \pi(y_{k+1}|z_{k+1}) \pi(z_{k+1}|y_{1:k}) dz_{k+1} \quad (5)$$

is the marginal density function. The likelihood function follows the same probability distribution as the random measurement error due to the mutually independency assumption.

Assuming a Gaussian initial state density distribution $\pi(z_0)$, and under the assumptions stated above, the updating formulas yield the Kalman filters (Kalman, 1960). Given $\pi(z_0) \propto \mathcal{N}(z_{0|0}, \Gamma_{0|0})$, the posterior filters follow a Gaussian distribution $\pi(z_k|y_{1:k}) \propto \mathcal{N}(z_{k|k}, \Gamma_{k|k})$ with mean $z_{k|l} = E[z_k|y_{1:l}]$ and covariance $\Gamma_{k|l} = \text{cov}(z_k|y_{1:l})$ that can be estimated as follows. The prediction density in the time-evolution updating in Eq. (3) is $\pi(z_{k+1}|y_{1:k}) \propto \mathcal{N}(z_{k+1|k}, \Gamma_{k+1|k})$ with

$$z_{k+1|k} = \mathbf{G}(\boldsymbol{\theta})z_{k|k} + F_k, \quad (6a)$$

$$\Gamma_{k+1|k} = \mathbf{G}(\boldsymbol{\theta})\Gamma_{k|k}\mathbf{G}(\boldsymbol{\theta})^T + \Gamma_w. \quad (6b)$$

The observation update in Eq. (4) corresponds to $\pi(z_{k+1}|y_{1:k+1}) \propto \mathcal{N}(z_{k+1|k+1}, \Gamma_{k+1|k+1})$ with

$$z_{k+1|k+1} = z_{k+1|k} + \mathbf{K}_{k+1}e_{k+1}, \quad (7a)$$

$$\Gamma_{k+1|k+1} = (1 - \mathbf{K}_{k+1}\mathbf{H})\Gamma_{k+1|k} \quad (7b)$$

where

$$e_{k+1} = y_{k+1} - \mathbf{H}z_{k+1|k} \quad (8)$$

is related to the likelihood, and K_{k+1} is the Kalman gain matrix

$$\begin{aligned} \mathbf{K}_{k+1} &= \Gamma_{k+1|k} \mathbf{H}^T \Gamma_{e_{k+1}}^{-1}, \\ \Gamma_{e_{k+1}} &= \mathbf{H} \Gamma_{k+1|k} \mathbf{H}^T + \Gamma_v \end{aligned} \quad (9)$$

for $k \geq 0$. These filters provide point estimates of the unknown state vectors dependent on the parameters $\boldsymbol{\theta}$ and conditioned to the observations and to the prior knowledge on the time-varying system. Furthermore, these formulae constitute the starting point for the parameter estimation algorithm described latter in this paper. The method proposed here is presented as an alternative to sampling-based methods which are not recommended for large dimensional problems due to unfeasible computational times. This is the case of the time evolution vocal fold model used in this work, derived from continuous stochastic differential equations. The evolution model is summarized in what follows.

2.1.1. Vocal fold evolution model

Deterministic model. The layered three-dimensional continuum vocal fold model is based on (Granados et al., 2016). Under the small displacement assumption, the deformation $\mathbf{u} \in \mathbb{R}^3$ of a vocal fold particle from its original position $\bar{\mathbf{x}} \in V_{\text{solid}} \subset \mathbb{R}^3$ to its current position $\mathbf{x} \in v_{\text{solid}} \subset \mathbb{R}^3$ can be modeled as

$$\nabla \cdot \boldsymbol{\sigma}(\boldsymbol{\theta}) = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \quad \text{in } v_{\text{solid}}, \quad (10a)$$

$$\mathbf{u} = \mathbf{0} \quad \text{on } \Gamma_D, \quad (10b)$$

$$\boldsymbol{\sigma} \cdot \mathbf{n} = \mathbf{p} \quad \text{on } \Gamma_N, \quad (10c)$$

where ρ is the spatial mass density and $\boldsymbol{\sigma}$ is the Cauchy stress tensor which describes a linear stress-strain viscoelastic constitutive law defined by a set of mechanical parameters $\boldsymbol{\theta}$, such as viscosity, Young's or shear modulus. The tissue-cricoid cartilage attachment region $\Gamma_D \subset \partial v_{\text{solid}}$ in Fig. 1(b) is modeled as Dirichlet boundary conditions with zero displacement in Eq. (10b). Furthermore, at the moving mucosal boundary $\Gamma_N \subset \partial v_{\text{solid}}$ in Eq. (10c), Neumann boundary condition are enforced. The aerodynamic forces \mathbf{p} are derived from a one-dimensional incompressible Bernoulli flow along a spatially discretized glottal canal into cross-sectional areas $a(z_i)$, assuming no vocal tract coupling and flow detachment point at minimum area. The transversal areas $a(z_i)$ are computed by assuming a constant portion of a circular sector and left and right semi-ellipses defined by the fixed most anterior and posterior points and the current medio-lateral components $x_{i,>0}$ and $x_{i,<0}$ of the medial point; see dashed line ($t > 0$) in Fig. 1(b). In addition, when vocal fold collision occurs, additional inequality boundary conditions on the contact surface $\Gamma_C \subset \partial v_{\text{solid}}$ arise to avoid body interpenetration (Granados et al., 2016). This condition make the problem non-linear and non-smooth, which increases the complexity of the statistical nonstationary inverse problem. Furthermore, high-speed imaging fails at extracting the motion at collision (Granados and Brunskog, 2015). Because simplifications may lead to poor estimates (Hadwin et al., 2016), it is here recommended to avoid including collision in the vocal fold model, and extract collision-free data by producing high-pitch sustained oscillations during high-speed laryngoscopy. However, proper mathematical formulations (e.g., see Shimada et al., 1998; Simon and Simon, 2006, 2010) may possibly be adapted to the present large problem.

A finite element discretization of the solid domain into n_{np} nodal points yields the matrix form of the continuum problem in Eq. (10)

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}(\boldsymbol{\theta})\dot{\mathbf{u}} + \mathbf{K}(\boldsymbol{\theta})\mathbf{u} = \mathbf{F} \quad (11)$$

where \mathbf{M} , \mathbf{C} , and \mathbf{K} are the mass, damping, and stiffness $3n_{np} \times 3n_{np}$ matrices, respectively, \mathbf{F} is the $3n_{np}$ -vector of applied aerodynamic forces, and \mathbf{u} is the $3n_{np}$ -vector that contain the displacement field at each node; dots indicate time derivatives, and for notational convenience no difference is made between particle and vector node deformation. The time discretization scheme chosen in this work is an implicit one-step

α -scheme, unconditionally stable and highly diffusive (Granados et al., 2016). The resulting system matrix is

$$\mathbf{A} \equiv \mathbf{A}(\boldsymbol{\theta}) = \mathbf{M} + (1 + \alpha)\gamma h \mathbf{C}(\boldsymbol{\theta}) + (1 + \alpha)h^2 \beta \mathbf{K}(\boldsymbol{\theta}) \quad (12)$$

with h the finite element time step and β , γ and α parameters of the time-integration scheme. The importance of the choice of one-step temporal integration schemes is discussed in the following section.

Stochastic model. Given a condensed form $\mathbf{A}_c(\boldsymbol{\theta})$ of the system matrix in Eq. (12), it is possible to write the discrete unconstrained problem in Eq. (11) as a state evolution model in Eq. (1a). The solution corresponds to the stochastic processes $\{U_k\}_{k=0}^N$, $\{\dot{U}_k\}_{k=0}^N$, $\{\ddot{U}_k\}_{k=0}^N$ of $3n_{np}$ -dimensional random variables, condensed in a state random vector $Z_k = \begin{pmatrix} U_k & \dot{U}_k & \ddot{U}_k \end{pmatrix}^T \in \mathbb{R}^{9n_{np}}$. For the sake of brevity, the full expression of the evolution matrix and the input vector F_k are presented in Appendix A. It must be emphasized, however, that the matrix contains non-linear components in $\boldsymbol{\theta}$, which is a key observation for the development of a parameter estimation method. In addition, the input vector F_k is composed of aerodynamic forces, which in turn depend on the intraglottal boundary defined by the current configuration. For the sake of mathematical simplicity, the vector is precomputed from the estimated vocal fold configuration at the previous time step, and explicit dependency is disregarded.

Furthermore, the Dirichlet equality constraint in Eq. (10b) can be enforced by a Lagrangian approach as in (Granados et al., 2016) leading to constrained Kalman filters (Simon and Chia, 2002). However, at least two drawbacks must be emphasized. Firstly, the Lagrange multiplier method finds the unknown boundary reaction forces to solve the constrained problem in an exact way, thereby increasing the number of unknown variables. Hence, the dimension of the evolution matrix $\mathbf{G}(\boldsymbol{\theta})$ is also increased. Despite the sparse nature of the matrix, it is recommended to keep the dimension as low as possible if significant matrix manipulation is required, as it occurs with the Kalman filters in Eqs. (6)-(9). Furthermore, the same argumentation supports the necessity of one-step temporal discretization schemes. A discrete stochastic process obtained from multi-step schemes is modeled as a high-order Markov chain, which leads to an observation matrix of large dimension (Kaipio and Somersalo, 2005). Secondly, the resulting system matrix of the Lagrangian solution is indefinite, which requires special numerical treatment. Alternatively, Dirichlet boundary conditions can be easily incorporated in the observation model by augmenting the measurement vector and assuming a noise-free measurement as in (Wang et al., 2002). This procedure is described in the next section.

The initial probability density function is assumed to be Gaussian, as previously stated, zero mean $z_{0|0}$, with diagonal covariance matrix $\Gamma_0 = \Gamma_{0|0}$ for computational convenience. The variances are large to account for the large uncertainty of the actual vocal fold motion. Similarly, the noise covariance Γ_v is assumed to be diagonal and time-independent, although an expression can be derived from algebraic manipulation of the observation model. However, as will be seen later in Sec. 2.2, it is possible to find estimates of the initial and noise distribution parameters.

2.1.2. Laryngeal observation model

Deterministic model. A non-invasive method to extract superior two-dimensional laryngeal dynamics is the optical flow (Granados and Brunskog, 2015), which provides dense estimates of the projected two-dimensional velocity $\mathbf{q}^s = (q_1^s, q_2^s)$ of three-dimensional moving objects captured in an image sequence. The algorithm is derived from the main assumption that brightness remains constant between time frames. Let a pixel located at a position $\mathbf{x} \in \mathbb{R}^2$ have brightness $I(\mathbf{x}, k) \in \mathbb{Z}^+$ at time step $k \in [1, N]$. It can be shown that the robust instant total optical flow $\mathbf{q} \in \mathbb{R}^{2n_{of}}$ within an image region $\{\mathbf{x}^s\}_{s \in \mathcal{S}}$ of n_{of} pixels indexed by \mathcal{S} is the solution to the regularization problem

$$\min_{\mathbf{q}} E(\mathbf{q}) = \min_{\mathbf{q}} \{E_D(\mathbf{q}) + \lambda E_S(\mathbf{q})\}, \quad (13)$$

where $\lambda \geq 0$ is the regularization parameter, and

$$E_D(\mathbf{q}) = \sum_{s \in \mathcal{J}} \rho_D(\nabla I(\mathbf{x}^s, k) \cdot \mathbf{q}^s + I_k(\mathbf{x}^s, k)), \quad (14)$$

$$E_S(\mathbf{q}) = \sum_{s \in \mathcal{J}} (\rho_S(\nabla q_1^s) + \rho_S(\nabla q_2^s)), \quad (15)$$

with the nabla operator being the spatial image gradient and $\rho_D = \rho_S = \rho$ a robust non-convex Lorentzian function (Black and Anandan, 1996).

Stochastic model. Assume that the pixel size can be estimated. The interface superior n'_{np} finite element points indexed by \mathcal{J}' could be mapped to a number of pixel positions via shape functions defined on the superficial triangular finite element, if pixel connectivities were defined at the initial step. However, because the optical flow algorithm may give local poor estimates (Granados and Brunsog, 2015), pixel advection further increases the errors in time leading to nonsense results. As an alternative, at each time-evolution step the closes pixel to each of the superficial finite element nodes can be found in a pre-processing step, and stored in a $2n'_{np}$ -dimensional vector according to the labeling of the superficial finite element mesh. Note that typically $n'_{np} \ll n_{of}$. Assume, then, that $q_k \in \mathbb{R}^{2n'_{np}}$ is the realization at frame k of this reduced observable optical flow and Q_k the corresponding random variable. Then, an observation matrix as in Eq. (1b) can be defined as $\mathbf{H}^{\mathbf{q}} = \{\mathbf{H}_{ij}^{\mathbf{q}}\}$ with

$$\mathbf{H}_{ij}^{\mathbf{q}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{if } j \in \mathcal{J}' \quad (16)$$

and null matrix otherwise, for $i = 1, \dots, n'_{np}$ and $j = 1, \dots, n_{np}$. If Dirichlet conditions are enforced by augmenting the measurement vector by a vector of zero displacements $U_k^D = 0$ at the nodes indexed by \mathcal{D} on Γ_D , by defining the measurement random vector as $Y_k = (Q_k \ U_k^D)^T$, the observation model in Eq. (1b) yields

$$Y_k = \begin{bmatrix} \mathbf{H}^{\mathbf{q}} & 0 & 0 \\ \mathbf{H}^D & 0 & 0 \end{bmatrix} \cdot Z_k + \begin{bmatrix} V \\ 0 \end{bmatrix}. \quad (17)$$

where $\mathbf{H}^D \in \mathbb{R}^{3|\mathcal{D}| \times 3n_{np}}$ corresponds to the finite element Dirichlet matrix of connectivities. Furthermore, the measurement noise is here modeled as a simple Gaussian with zero mean and a diagonal covariance matrix. Similarly to the estimates in the previous evolution model, it is possible to find estimates of the noise.

From what presented here, this method requires that the time step between consecutive observation is equal to the time step between model states. Endoscopic high-speed imaging techniques typically record at about 2000 – 4000 frames per second, which corresponds to frame steps about 500 – 250 μs . Even though unconditionally stable temporal discretization schemes provide stable solutions at large time steps, the numerical error in Eq. (11) can be large (Granados et al., 2016). Acceptable state time steps correspond to unfeasible laryngoscopic frame rate about 10000 frames per second. However, small evolution steps can be used by considering an state-space model with missing observations. In such a situation, specific maximum-likelihood algorithms for parameter estimation are designed to handle state-spaces step mismatch (Digalakis et al., 1993). This algorithm is described as follows.

2.2. State-space identification problem

The problem of estimating the parameters θ based on observations and prior knowledge of the time-varying system is known as state-space identification problem (Kaipio and Somersalo, 2005) or Bayesian model updating (Ching and Chen, 2007; Simoen et al., 2013). Typically, maximum-likelihood techniques (Digalakis et al., 1993) or particle filters methods (Doucet and Tadić, 2003; Ching and Chen, 2007; Hadwin et al., 2016) are employed. The use of the latter methods is appropriate for a general non-linear and non-Gaussian inverse problem. In that case, the solution takes the form of a sampled conditional probability distribution, rather than single-valued point estimates as with the Kalman filters. However, large sample

generation is typically required, which makes the method computationally unfeasible for finite element models such as the present one. For this reason, the linear and Gaussian assumption is made in the present continuum vocal fold problem. However, the linearity with respect the state vector in Eq. (1b) is not longer valid if the unknown mechanical parameters θ are also modeled as a time-varying stochastic process $\{\Theta_k\}_{k \geq 0}$ instead of static unknown system parameters. In that case, the parameter random variable would be included in an augmented state vector $(Z_k \ \Theta_k)^T$ (Ching et al., 2006) satisfying a certain evolution model, thereby a non-linear system would result; see the evolution matrix in Appendix A. Extended Kalman filters could then be used to extract point estimates as an alternative to particle filters, by applying Kalman filtering to the linearized problem (Kaipio and Somersalo, 2005). Thus, finding an expression of the Jacobian of the evolution matrix is straightforward if Θ models the viscosity, but complex expression may occur if elastic parameters are wished to be estimated in a finite element context. Therefore, the mechanical parameters are assumed to be time-independent and the identification problem is solve with a maximum-likelihood algorithm suitable for missing observation data, which can be summarized as follows.

Consider the set of observation $y_{1:N} = \{y_1, \dots, y_N\}$, and a parameter vector ξ ; a different notation is used for the parameter vector for what will be seen latter in this section. A classical maximum-likelihood approach consists of solving the problem

$$\hat{\xi} = \arg \max_{\xi \in \Xi} \mathcal{L}(\xi; y_{1:N}) = \arg \max_{\xi \in \Xi} \log \pi(y_{1:N}; \xi) \quad (18)$$

in the parameter space Ξ . The objective function can be seen as a marginal log-likelihood function as it considers the observations regardless of the underlying system, as in Eq. (5). However, these calculations can be exceedingly cumbersome. Alternatively, the log-likelihood can be defined as the conditional (incomplete) log-likelihood

$$\mathcal{L}(\xi; y_{1:N} | z_{0:N}) = \sum_{k=1}^N \log \pi(y_k | z_k; \xi) = -\frac{1}{2} \left(\sum_{k=1}^N \log |\Gamma_{e_k}| + \sum_{k=1}^N e_k^T \Gamma_{e_k}^{-1} e_k \right) \quad (19)$$

where the Kalman filters in Eqs. (6)- (9) have been used in the second equality, which implicitly depend on the biomechanical parameters. By taking the negative of the log-likelihood and setting $\xi = \theta$, numerical techniques can be applied to find a maximum-likelihood biomechanical parameter estimates. However, the performance of the algorithm is highly sensitive to the nature of the system matrices and the choice of the initial estimate, which makes the approach not optimal for the present finite element model. In addition, as stated earlier in this paper, the vocal fold state-space model may contain missing observations, which requires complex modifications of the Eq. (19). This suggests the use of the following expectation-maximization (Dempster et al., 1977).

Besides the observations $y_{1:N}$, consider the history of the state vector $z_{0:N} = \{z_0, \dots, z_N\}$ and the input data samples $f_{0:N} = \{f_0, \dots, f_N\}$. The problem is now to find the parameter vector that maximizes the joint probability distribution evaluated at the complete data outcome of the state-space model $\pi(y_{1:N}, z_{0:N}; \xi)$. That is, find a maximizer in Ξ of the complete log-likelihood (Ching et al., 2006)

$$\begin{aligned} \mathcal{L}(\xi; y_{1:N}, z_{0:N}) &= \log \pi(y_{1:N}, z_{0:N}; \xi) = \mathcal{L}(\xi; y_{1:N} | z_{0:N}) + \log \pi(z_{0:N}; \xi) \\ &= \sum_{k=1}^N \log \pi(y_k | z_k; \xi) + \log \pi(z_0; \xi) + \sum_{k=1}^N \log \pi(z_k | z_{k-1}; \xi). \end{aligned} \quad (20)$$

This expression can be easily derived from Bayes' formula and from the fact that $\pi(y_k | z_{0:k}, y_{1:k-1}) = \pi(y_k | z_k)$, because $\{Y_k\}_{k \geq 1}$ is a Markov process with respect to the history of $\{Z_k\}_{k \geq 0}$, as stated in Sec. 2.1. Assume further that ξ^t is an initial guess of the maximizer. Then, it can be shown that the inequality

$$\mathcal{L}(\xi; y_{1:N}) - \mathcal{L}(\xi^t; y_{1:N}) \geq Q(\xi; \xi^t) - Q(\xi^t, \xi^t) \quad (21)$$

holds, with

$$Q(\xi, \xi^t) = E_{\xi^t} \{ \mathcal{L}(\xi; y_{1:N}, z_{0:N}) | y_{1:N} \} = \int \mathcal{L}(\xi; y_{1:N}, z_{0:N}) \pi(z_{0:N} | y_{1:N}; \xi^t) dz_{0:N} \quad (\text{E-step}) \quad (22a)$$

$$\xi^{t+1} = \arg \max_{\xi \in \Xi} Q(\xi, \xi^t) \quad (\text{M-step}), \quad (22b)$$

which corresponds to the so-called expectation (E) and maximization (M) steps, respectively. The inequality and the recursive schema assures that ξ^{t+1} is a good approximation of the maximum-likelihood estimate $\hat{\xi}$ of the marginal log-likelihood in Eq. (18) when convergence is reached. Assuming a linear and Gaussian problem, the Kalman filters can be used to derive a simple expression of the complete log-likelihood, as with Eq. (19). Then, taking conditional expectations, the Q -function in Eq. (22a) yields (Shumway and Stoffer, 1982)

$$\begin{aligned} Q(\xi, \xi^t) = & -\frac{1}{2} \log |\Gamma_0| - \frac{N}{2} \log |\Gamma_w| |\Gamma_v| - \frac{1}{2} Tr \left\{ \Gamma_0^{-1} \cdot E_{\xi^t} \{ (z_0 - z_{0|0})(z_0 - z_{0|0})^T | y_{0:N} \} \right. \\ & - \frac{1}{2} Tr \left\{ \Gamma_w^{-1} \cdot \left[\Sigma_1 - \Sigma_2 \cdot (\mathbf{G}(\theta) \ \mathbf{I})^T - (\mathbf{G}(\theta) \ \mathbf{I}) \cdot \Sigma_2^T + (\mathbf{G}(\theta) \ \mathbf{I}) \cdot \Sigma_3 \cdot (\mathbf{G}(\theta) \ \mathbf{I})^T \right] \right\} \\ & \left. - \frac{1}{2} Tr \left\{ \Gamma_v^{-1} \cdot \left[\Sigma_4 - \Sigma_5 \mathbf{H}^T - \mathbf{H} \Sigma_5^T - \mathbf{H} \Sigma_1 \mathbf{H}^T \right] \right\} \right\} \quad (23) \end{aligned}$$

where $Tr\{\cdot\}$ indicates the trace of the matrix and

$$\begin{aligned} \Sigma_1 = \sum_{k=1}^N E_{\xi^t} \{ z_{k-1} z_{k-1}^T | y_{1:N} \}, \quad \Sigma_2 = \sum_{k=1}^N \left(E_{\xi^t} \{ z_k z_{k-1}^T | y_{1:N} \} E_{\xi^t} \{ z_k f_{k-1}^T | y_{1:N} \} \right) \\ \Sigma_3 = \sum_{k=1}^N \left(\begin{matrix} E_{\xi^t} \{ z_{k-1} z_{k-1}^T | y_{1:N} \} & E_{\xi^t} \{ z_{k-1} f_{k-1}^T | y_{1:N} \} \\ E_{\xi^t} \{ f_{k-1} z_{k-1}^T | y_{1:N} \} & E_{\xi^t} \{ f_{k-1} f_{k-1}^T | y_{1:N} \} \end{matrix} \right), \quad \Sigma_4 = \sum_{k=1}^N E_{\xi^t} \{ y_k y_k^T | y_{1:N} \}, \quad \Sigma_5 = \sum_{k=1}^N E_{\xi^t} \{ y_k z_k^T | y_{1:N} \}. \end{aligned} \quad (24)$$

These statistics can be calculated, even for missing observations (Digalakis et al., 1993). Making use of Bayesian smoothing filters in a backward iteration (Digalakis et al., 1993; Kaipio and Somersalo, 2005), the calculations in Eq. (24) require

$$\begin{aligned} E_{\xi^t} \{ y_k z_k^T | y_{1:N} \} &= \begin{cases} y_k E_{\xi^t} \{ z_k | y_{1:N} \} & \text{if observed} \\ \mathbf{H} E_{\xi^t} \{ z_k z_k^T | y_{1:N} \} & \text{if missing} \end{cases} \\ E_{\xi^t} \{ y_k y_k^T | y_{1:N} \} &= \begin{cases} y_k y_k^T & \text{if observed} \\ \Gamma_v + \mathbf{H} E_{\xi^t} \{ z_k z_k^T | y_{1:N} \} \mathbf{H}^T & \text{if missing} \end{cases} \\ E_{\xi^t} \{ z_k z_k^T | y_{1:N} \} &= z_{k|N} z_{k|N}^T + \Gamma_{k|N} \\ E_{\xi^t} \{ z_k z_{k-1}^T | y_{1:N} \} &= z_{k|N} z_{k-1|N}^T + \Gamma_{k,k-1|N} \\ E_{\xi^t} \{ y_k | y_{1:N} \} &= \begin{cases} y_k & \text{if observed} \\ \mathbf{H} E_{\xi^t} \{ z_k | y_{1:N} \} & \text{if missing.} \end{cases} \end{aligned} \quad (25)$$

Define now the parameter vector

$$\xi = \left((\mathbf{G}(\theta) \ \mathbf{I}) \ \mathbf{H} \ z_{0|0} \ \Gamma_{0|0} \ \Gamma_w \ \Gamma_v \right), \quad (26)$$

containing the system matrix and the parameters of the random distributions, which have been assumed to be known earlier in this paper, even though the actual value is highly uncertain. This parameter vector implicitly depend on the vocal fold biomechanical parameters θ , to be estimated in this work. It can be seen that the maximum ξ^{t+1} at the M-step in Eq.(22b) occurs at

$$\xi^{t+1} = \left((\Sigma_2 \Sigma_3^{-1}) \ \Sigma_5 \Sigma_1^{-1} \ z_{0|N} \ \Gamma_{0|N} \ \frac{1}{N} (\Sigma_1 - \Sigma_2 \Sigma_3^{-1} \Sigma_2^T) \ \frac{1}{N} (\Sigma_4 - \Sigma_5 \Sigma_1^{-1} \Sigma_5^T) \right). \quad (27)$$

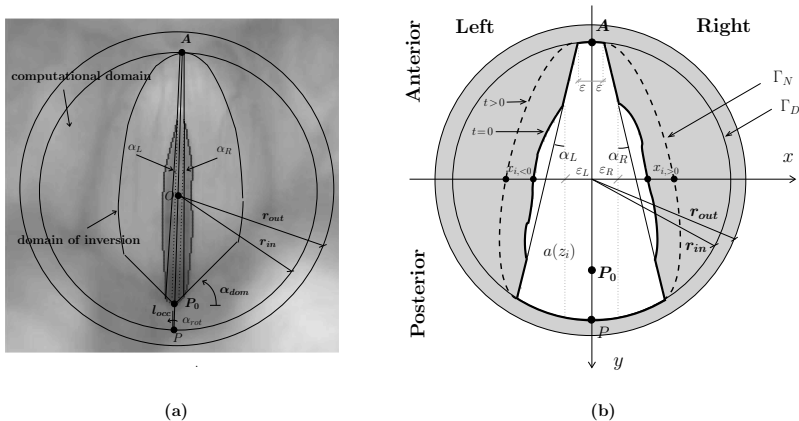


Figure 1: Superior view of the vocal folds. (a) Intraglottal area segmentation and geometry definition of the prephonatory frame of a high-speed laryngoscopic image sequence, before rotation. (b) Schematic representation of the initial cross-sectional evolution model geometry. Dashed line indicates the assumed tissue deformation for Bernoulli pressure computations, bounded by the medial medio-lateral nodal components $x_{i,<0}$ and $x_{i,>0}$. User-defined input parameters (in bold): r_{in} , r_{out} , l_{occ} , A and P_0 .

The E-M steps in Eq. (22) are repeated iteratively. At termination, estimates of the probability distribution of the state and measurement noise, as well as the initial state are obtained. Finally, the mechanical continuum parameters θ may be calculated from the final estimated evolution matrix $\mathbf{G}(\theta)$.

2.3. Model generation framework

The state-space model developed in this work requires an automatic finite element mesh generation method based on the prephonatory vocal fold image. If the initial transient is not included in the image measurement sequence, a time step must be selected to match the observation model. However, this arbitrary choice carries a large uncertainty, which may have severe consequences in the correct formulation of the state-space model. Furthermore, it is worth pointing out that the initial choice of the mechanical parameters θ would most likely lead to a fundamental frequency of vibration different from the actual frequency captured by the image sequence. However, in that case the measurement error V would account for the frequency difference as an amplitude error, thereby avoiding the sample mismatch problem. Furthermore, a fine cross-sectional initial mesh is necessary to build the observation model with certain accuracy. The large number of degrees of freedom lead to a large evolution matrix, and hence further implementation analysis is required to handle the complex numerical manipulation of Kalman filters.

An image segmentation algorithm for edge-detection based on scale-space representation (Lindeberg, 1996) is used to estimate the cross-sectional intraglottal finite element initial geometry. Before the segmentation algorithm is applied, the intraglottal midline is manually defined by selecting the most anterior point A (anterior commissure) and posterior point P_0 . Then, a rotation α_{rot} is applied. This is illustrated in Fig. 1(a), as well as the result of the intraglottal region segmentation in light gray. In this section bold text indicates manually selected points and parameters. Furthermore, the anterior part of the vocal folds may be in contact due to the small anterior-posterior initial strain related to intrinsic muscles relaxation. To avoid contact modeling, the initial configuration is approximated by a circular sector in the anterior and posterior regions with α_R , α_L and ϵ obtained by simple trigonometry; see Fig. 1(b) for an illustration. Notice that the circular sector between P_0 and P (posterior commissure) in Fig. 1(b) is occluded in the image frame in Fig. 1(a) by the corniculate cartilage. The occluded geometry would ideally be estimated by statistical shape analysis if an vocal fold atlas existed (Cootes et al., 1995). Alternatively, an occluded vocal fold length l_{occ} and α_{dom} must be predefined based on previous studies on the arytenoid cartilage geometry.

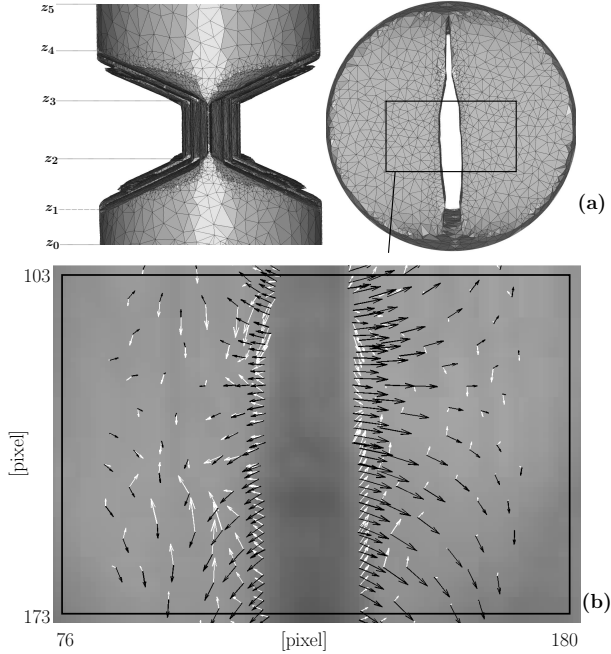


Figure 2: (a) Three-dimensional initial finite element mesh, with manually selected input parameters z_0 , z_1 , z_2 , z_3 , z_4 and z_5 . (b) Black arrows indicate the displacement field U_k at a number of superior finite element nodes \mathcal{S}' within a selected region of the domain of inversion (top, right) in the image frame of reference; white arrows indicate the optical flow at the pixels closes to the node position. Results at $t = 0.0065 \mu s$

Hence the domain of inversion containing the n'_{np} finite element nodal points can be defined. Finally, to define the state-observation mapping in the observation model in Eq. (17), it is necessary to estimate the pixel size. However, accurate estimates cannot be made from mono images of the superior surface motion recorded with a high-speed camera. Similarly to the occluded length, the total length $2\mathbf{r}_{in}$ and the larynx external boundary $\mathbf{r}_{out} - \mathbf{r}_{in}$ must be predefined, and thereby the pixel size can be estimated. Finally, in regard to the three-dimensional geometry, the inferior-superior points z_0 , z_1 , z_2 , z_3 , z_4 and z_5 at the left of Fig. 2(a) must be predefined, as well as the thickness of each layer. Then, a finite element mesh can be built in a similar way as in (Granados et al., 2016). It is worth mentioning that the three-dimensional geometry is arbitrary, and further studies must be carried out to investigate the effect on the tissue deformation, specially within the region compressed between z_2 and z_4 .

3. Simulations setup and discussion

High-speed movies of the human vocal fold vibration including initial transient and steady vibrations are recorded from a female voice-healthy subject with a digital camera (HRES Endocam, model 5562.9 color; Wolf, Germany) at a frame rate of 4000 frames per second. The image resolution is 254×254 . The initial frame at the phonation onset is converted to a grayscale image with with a MATLAB routine according to the transformation $0.2989R + 0.5870G + 0.1140B$. A custom finite element code is used for the simulations based on the work by (Granados et al., 2016).

The tetrahedral finite element mesh is generated with the Computational Geometry Algorithm Library (Alliez et al., 2015) from the initial frame in Fig. 1(a) as described in the previous section. The geometrical

Table 1: Tissue parameters θ used in the finite element simulation for a normal larynx as in (Granados et al., 2016). (E, ν) , transversal Young’s modulus and Poisson’s ratio; (E', μ', ν') , longitudinal Young’s and shear modulus, and Poisson’s ratio. η indicates viscosity. All values are specified in the centimetre-gram-second system of units.

	E	ν	E'	μ'	ν'	η
Cover	$1 \cdot 10^4$	0.66	$10 \cdot 10^4$	$10 \cdot 10^4$	0.66	4
Ligament	$3 \cdot 10^4$	0.58	$20 \cdot 10^4$	$40 \cdot 10^4$	0.58	10
Body	$20 \cdot 10^4$	0.45	$40 \cdot 10^4$	$30 \cdot 10^4$	0.45	10

input parameters are $l_{occ} = 0.1$ cm, $r_{in} = 0.7$ cm, $r_{out} = 0.8$ cm, $z_0 = -1.2$ cm, $z_1 = -0.7$ cm, $z_2 = -0.4$ cm, $z_3 = 0$ cm, $z_4 = 0.2$ cm, and $z_5 = 1.2$ cm; see Fig. 2(a). A subglottal pressure of 0.8 kPa and the tissue parameters θ shown in Table 1 are used for the three-layered vocal fold finite element model in Fig. 2(a), with predefined layers thickness of 0.05 cm. Furthermore, the optical flow algorithm is applied to the entire image domain, with $\lambda = 0.5$ in Eq. (13) (Granados and Brunskog, 2015). The domain of inversion is shown in Fig. 1(a) for $\alpha_{dom} = 45$ degrees and an ellipse with major axis $0.5|\overline{AP}|$, minor axis $0.4r_{in}$ centered at O . It is worth mentioning, that the finite element model contains more than 30000 degrees of freedom. This observation raises the problem of large matrix manipulation in the Kalman equations. The complexity of the matrix manipulations could be drastically decreased if sparsity was assumed. Despite the sparse nature of the finite element matrix $\mathbf{G}(\theta)$ and the observation matrix \mathbf{H} , the mean and covariance Kalman estimates cannot be assumed to be sparse. In addition, inverting matrices, in particular covariance matrices, may require special numerical attention. This is because covariance matrix estimates may be almost singular given the similarity of the samples. However, specific linear algebra libraries for matrix manipulation could be used with the custom finite element code (Granados et al., 2016) in order to solve the computational limitations. Further implementation is therefore required in order to evaluate the present inverse method.

In Fig. 2(b) an illustration of the observation model defined in Eq. (17) is shown, at a time step $t = 0.0065 \mu s$. In black arrows a realization u_k of the finite element random vector U_k at a number of superior finite element nodes in \mathcal{S}' is shown, within a selected region of the domain of inversion (Fig. 2(a), right) in the image frame of reference. In white arrows, the realization q_k of the random optical flow. From the illustration, it is noted that the selected tissue parameters θ shown in Table 1 generate a vibration pattern different from actual one captured by the optical flow. The selected time instant corresponds to an opening phase (note the intraglottal edge vectors) with a complex superficial tissue deformation: while the finite element model estimates positive strains in the middle region, pulling the tissue toward the anterior-posterior extremes, the optical flow captures the opposite behavior, indicating compression of the tissue. The expectation-maximization algorithm for parameter estimation is expected to correct for these differences, accounting for measurement and modeling errors. In addition, the expectation-maximization algorithm could be validated by limiting the unknown parameter vector θ to the viscosity. Assuming a stochastic time-dependent viscosity, the state vector could be augmented and in that case easy expressions of the extended Kalman filter could be found, as discussed earlier in Sec. 2.2.

Finally, once the computational limitations are overcome, the procedure could be validated with synthetic vocal folds (Murray and Thomson, 2012). Alternatively, synthetic vocal fold deformations generated by the finite element model and visualized with an advance three-dimensional illuminance software could be used for the validation.

4. Conclusions

A statistical formulation for vocal fold continuum tissue parameters estimation has been presented and discussed. The problem has been formulated as a linear and Gaussian state-space identification problem. The model based on robust optical flow data extraction and a finite element discretization has been shown to provide dense information of the deformation, which is expected to reduce the ill-posedness of the inverse problem and provide accurate parameter estimates. Bayesian filtering methods has been used to develop the maximum-likelihood based algorithm for single-valued biomechanical parameters estimation.

The expectation-maximization algorithm for missing observations has been proposed as a solution to possible sample mismatch between the finite element time step and the video frame rate. Furthermore, the use of a finite element model has posed the problem of large matrix manipulations, which is attenuated by modeling Dirichlet conditions as noise-free measurement. It has been recommended to capture laryngeal videos during sustained phonation free of collision, thereby reducing the observation error at collision times and the computational complexity of the problem. Finally, the necessity of capturing the initial transient in the image sequences has been discussed.

Acknowledgment

This work was supported by grant number 110230 from the Swedish organization AFA Försäkring. The authors would like to express their gratitude to Dr. Marek Krzysztof Misztal (Niels Bohr Institute, Copenhagen, Denmark) who made important contributions to the finite element mesh generation algorithm.

Appendix A.

Given a state vector $Z_k = (U_k \dot{U}_k \ddot{U}_k)^T \in \mathbb{R}^{9n_p}$ and a mechanical parameter vector θ , simple algebraic manipulations yields the finite-element evolution model

$$Z_{k+1} = \mathbf{G}(\theta)Z_k + F_k + W_k, \quad (\text{A.1})$$

with the input vector being

$$F_k = \mathbf{A}_c^{-1}(\theta) \cdot \begin{bmatrix} \beta^2 \mathbf{F} \\ \gamma h \mathbf{F} \\ \mathbf{F} \end{bmatrix} \in \mathbb{R}^{9n_p}, \quad (\text{A.2})$$

and the evolution matrix $\mathbf{G}(\theta) = \{\mathbf{G}_{ij}(\theta)\}_{i,j=1,\dots,3} \in \mathbb{R}^{9n_p \times 9n_p}$ composed of the $3n_p \times 3n_p$ -dimensional matrix blocks

$$\begin{aligned} \mathbf{G}_{11}(\theta) &= 1 + \beta^2 \mathbf{A}_c^{-1}(\theta) \\ \mathbf{G}_{12}(\theta) &= h - \beta h^2 [\mathbf{A}_c^{-1}(\theta)(\mathbf{K}(\theta)(1 + \alpha)h + \mathbf{C})] \\ \mathbf{G}_{13}(\theta) &= \frac{h^2}{2}(1 - 2\beta) - \beta^2 \left[\mathbf{A}_c^{-1}(\theta)\mathbf{K}(\theta)(1 + \alpha)\frac{h^2}{2}(1 - 2\beta) + \mathbf{C}(1 + \alpha)(1 - \gamma)h \right] \\ \mathbf{G}_{21}(\theta) &= \gamma h \mathbf{A}_c^{-1}(\theta) \\ \mathbf{G}_{22}(\theta) &= 1 - \gamma h \mathbf{A}_c^{-1}(\theta) [\mathbf{K}(\theta)(1 + \alpha)h + \mathbf{C}] \\ \mathbf{G}_{23}(\theta) &= (1 - \gamma)h - \gamma h \mathbf{A}_c^{-1}(\theta)\mathbf{K}(\theta)(1 + \alpha)\frac{h^2}{2}(1 - 2\beta) + \mathbf{C}(1 + \alpha)(1 - \gamma)h \\ \mathbf{G}_{31}(\theta) &= \mathbf{A}_c^{-1}(\theta) \\ \mathbf{G}_{32}(\theta) &= -\mathbf{A}_c^{-1}(\theta) [\mathbf{K}(\theta)(1 + \alpha)h + \mathbf{C}] \\ \mathbf{G}_{33}(\theta) &= -\mathbf{A}_c^{-1}(\theta)\mathbf{K}(\theta)(1 + \alpha)\frac{h^2}{2}(1 - 2\beta) + \mathbf{C}(1 + \alpha)(1 - \gamma)h, \end{aligned} \quad (\text{A.3})$$

where h is the model time step, β , γ and α are the parameters of the implicit one-step α -scheme, and $\mathbf{A}_c(\theta)$ the condensed finite element system matrix $\mathbf{A}(\theta)$ (Granados et al., 2016).

References

Alliez, P., Jamin, C., Rineau, L., Tayeb, S., Tournois, J., Yvinec, M., 2015. 3D Mesh Generation, 4th Edition. CGAL Editorial Board.

- Andrade-Miranda, G., Henrich Bernardoni, N., Godino-Llorente, J. I., 2015. Optical-flow kymograms and glottovibrograms : a new way to present high-speed data for laryngeal assessment. In: 9th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Application. Firenze, Italy, pp. 71–74.
- Black, M. J., Anandan, P., 1996. The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields. *Comput. Vis. Image Und.* 63 (1), 75–104.
- Cataldo, E., Soize, C., Sampaio, R., 2013. Uncertainty quantification of voice signal production mechanical model and experimental updating. *Mech. Syst. Signal Pr.* 40 (2), 718–726.
- Ching, J., Beck, J. L., Porter, K. A., 2006. Bayesian state and parameter estimation of uncertain dynamical systems. *Probabilistic Eng. Mech.* 21 (1), 81–96.
- Ching, J., Chen, Y.-C., 2007. Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging. *J. Eng. Mech.* 133 (7), 816–832.
- Cootes, T., Taylor, C., Cooper, D., Graham, J., 1995. Active Shape Models-Their Training and Application. *Comput. Vis. Image Und.* 61 (1), 38–59.
- Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* 39 (1), 1–38.
- Digalakis, V., Rohlicek, J., Ostendorf, M., 1993. ML estimation of a stochastic linear system with the EM algorithm and its application to speech recognition. *IEEE Speech Audio Process.* 1 (4), 431–442.
- Döllinger, M., Hoppe, U., Hettlich, F., Lohscheller, J., Schuberth, S., Eysholdt, U., aug 2002. Vibration parameter extraction from endoscopic image series of the vocal folds. *IEEE Trans. Biomed. Eng.* 49 (8), 773–781.
- Doucet, A., Tadić, V. B., 2003. Parameter estimation in general state-space models using particle methods. In: *Ann. Inst. Stat. Math.* Vol. 55. pp. 409–422.
- Granados, A., Brunskog, J., 2015. Inverse problem in high-speed recordings of the vocal folds. In: *Book of Abstracts of the 11th Pan-European Voice Conference*. Florence, Italy.
- Granados, A., Misztal, M. K., Brunskog, J., Visseq, V., Erleben, K., 2016. A numerical strategy for finite element modeling of frictionless asymmetric vocal fold collision. *Int. J. Numer. Meth. Biomed. Engng.*
- Hadwin, P. J., Galindo, G. E., Daun, K. J., Zañartu, M., Erath, B. D., Cataldo, E., Peterson, S. D., 2016. Non-stationary Bayesian estimation of parameters from a body cover model of the vocal folds. *J. Acoust. Soc. Am.* 139 (5), 2683–2696.
- Hamilton, J. D., 1994. *Time Series Analysis*, 1st Edition. Princeton University Press, Princeton.
- Kaipio, J., Somersalo, E., 2005. *Statistical and Computational Inverse Problems*. Springer-Verlag New York.
- Kalman, R., 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.-T ASME* 82 (1), 35–45.
- Lindeberg, T., 1996. Scale-space: A framework for handling image structures at multiple scales. In: *Proc. CERN School of Computing*. Egmond aan Zee, The Netherlands.
- Misztal, M. K., Erleben, K., Bargteil, A., Fursund, J., Christensen, B. B., Andreas Bærentzen, J., Bridson, R., 2014. Multiphase flow of immiscible fluids on unstructured moving meshes. *IEEE Trans. Vis. Comput. Graphics* 20 (1), 4–16.
- Murray, P. R., Thomson, S. L., 2012. Vibratory responses of synthetic, self-oscillating vocal fold models. *J. Acoust. Soc. Am.* 132 (September), 3428–3438.
- Pinheiro, A. P., Stewart, D. E., MacIel, C. D., Pereira, J. C., Oliveira, S., 2012. Analysis of nonlinear dynamics of vocal folds using high-speed video observation and biomechanical modeling. *Digital Signal Process.* 22 (2), 304–313.
- Schmidt, B., Stingl, M., Leugering, G., Berry, D. a., Döllinger, M., 2011. Material parameter computation for multi-layered vocal fold models. *J. Acoust. Soc. Am.* 129 (4), 2168–80.
- Schwarz, R., Döllinger, M., Wurzbacher, T., Eysholdt, U., Lohscheller, J., 2008. Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model. *J. Acoust. Soc. Am.* 123 (5), 2717–2732.
- Shimada, N., Shirai, Y., Kuno, Y., Miura, J., 1998. Hand gesture estimation and model refinement using monocular camera-ambiguity limitation by inequality constraints. In: *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*. pp. 268–273.
- Shumway, R. H., Stoffer, D. S., 1982. An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.* 3 (4), 253–264.
- Simoen, E., Papadimitriou, C., Lombaert, G., 2013. On prediction error correlation in Bayesian model updating. *J. Sound Vibration* 332 (18), 4136–4152.
- Simon, D., Chia, T. L. I., 2002. Kalman filtering with state equality constraints. *IEEE Trans. Aerosp. Electron. Syst.* 38 (1), 128–136.
- Simon, D., Simon, D. L., 2006. Kalman filtering with inequality constraints for turbofan engine health estimation. In: *IEE Proc.-Control Theory Appl.* Vol. 153. pp. 371–378.
- Simon, D., Simon, D. L., 2010. Constrained Kalman filtering via density function truncation for turbofan engine health estimation. *Int. J. Syst. Sci.* 41 (2), 159–171.
- Spencer, M., Siegmund, T., Mongeau, L., feb 2008. Determination of superior surface strains and stresses, and vocal fold contact pressure in a synthetic larynx model using digital image correlation. *J. Acoust. Soc. Am.* 123 (2), 1089–1103.
- Steinecke, I., Herzog, H., 1995. Bifurcations in an asymmetric vocal-fold model. *J. Acoust. Soc. Am.* 97 (3), 1874–84.
- Wang, L.-S., Chiang, Y.-T., Chang, F.-R., 2002. Filtering method for nonlinear systems with constraints. *IEE Proc.-Control Theory Appl.* 149 (6), 525–531.
- Yang, A., Stingl, M., Berry, D. A., Lohscheller, J., Voigt, D., Eysholdt, U., Döllinger, M., 2011. Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model. *J. Acoust. Soc. Am.* 130 (2), 948–64.
- Zhang, Y., Jiang, J. J., 2008. Nonlinear dynamic mechanism of vocal tremor from voice analysis and model simulations. *J. Sound Vibration* 316 (1-5), 248–262.

Bibliography

- Åhlander, V. L., Rydell, R., and Löfqvist, A. (2011). Speaker's comfort in teaching environments: Voice problems in swedish teaching staff. *Journal of Voice*, 25(4):430–440.
- Alipour, F., Berry, D. A., and Titze, I. R. (2000). A finite-element model of vocal-fold vibration. *Journal of the Acoustical Society of America*, 108(6):3003–3012.
- Anandan, P. (1989). A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310.
- Bakhshaei, H. (2013). *Computational Models of Human and Animal Larynx and Vocal Folds*. McGill theses.
- Beck, J. L. (2010). Bayesian system identification based on probability logic. *Structural Control and Health Monitoring*, 17(7):825–847.
- Bernard, C. (1865). *An introduction to the study of experimental medicine*. Dover Publications.
- Berry, D., Montequin, D. W., and Tayama, N. (2001). High-speed digital imaging of the medial surface of the vocal folds. *Journal of the Acoustical Society of America*, 110(5):2539–2547.
- Berry, D., Zhang, Z., and Neubauer, J. (2006). Mechanisms of irregular vibration in a physical model of the vocal folds. *Journal of the Acoustical Society of America*, 120:EL36–EL42.

- Berry, D. A. and Titze, I. R. (1996). Normal modes in a continuum model of vocal fold tissues. *Journal of the Acoustical Society of America*, 100(5):3345–3354.
- Bertsekas, D. (1982). *Constrained optimization and Lagrange multiplier methods*. Academic Press.
- Black, M. J. and Anandan, P. (1996). The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields. *Computer Vision and Image Understanding*, 63(1):75–104.
- Bless, D., Hirano, M., and Feder, R. J. (1987). Videostroboscopic evaluation of the larynx. *Ear, Nose & Throat Journal*, 66:289–296.
- Cataldo, E., Soize, C., and Sampaio, R. (2013). Uncertainty quantification of voice signal production mechanical model and experimental updating. *Mechanical Systems and Signal Processing*, 40(2):718–726.
- Chan, R. W., Gray, S. D., and Titze, I. R. (2001). The importance of hyaluronic acid in vocal fold biomechanics. *Otolaryngology - Head and Neck Surgery*, 124(6):607–614.
- Chen, Z. H. E. (2003). Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond. *Statistics*, 182(1):1–69.
- Ching, J., Beck, J. L., and Porter, K. A. (2006). Bayesian state and parameter estimation of uncertain dynamical systems. *Probabilistic Engineering Mechanics*, 21(1):81–96.
- Ching, J. and Chen, Y.-C. (2007). Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging. *Journal of Engineering Mechanics*, 133(7):816–832.
- Chopin, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Annals of statistics*, pages 2385–2411.
- Cisonni, J., Van Hirtum, A., Pelorson, X., and Willems, J. (2008). Theoretical simulation and experimental validation of inverse quasi-one-dimensional steady and unsteady glottal flow models. *The Journal of the Acoustical Society of America*, 124(1):535–45.
- Cootes, T., Taylor, C., Cooper, D., and Graham, J. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, 61(1):38–59.

- Corigliano, A. and Mariani, S. (2004). Parameter identification in explicit structural dynamics: Performance of the extended Kalman filter. *Computer Methods in Applied Mechanics and Engineering*, 193(36-38):3807–3835.
- Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.
- Czermak, J. N. (1861). *On the laryngoscope, and its employment in physiology and medicine*. London, Englan: The New Sydenham Society.
- Deliyski, D. D., Petrushev, P. P., Bonilha, H. S., Gerlach, T. T., Martin-Harris, B., and Hillman, R. E. (2008). Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution. *Folia Phoniatica et Logopaedica*, 60(1):33–44.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.
- Döllinger, M., Hoppe, U., Hettlich, F., Lohscheller, J., Schuberth, S., and Eysholdt, U. (2002). Vibration parameter extraction from endoscopic image series of the vocal folds. *IEEE Transactions on Biomedical Engineering*, 49(8):773–781.
- Doucet, A. and Tadić, V. B. (2003). Parameter estimation in general state-space models using particle methods. In *Annals of the Institute of Statistical Mathematics*, volume 55, pages 409–422.
- Eftekhar Azam, S., Bagherinia, M., and Mariani, S. (2012). Stochastic system identification via particle and sigma-point Kalman filtering. *Iranian Journal of Science and Technology*, 19(4):982–991.
- Farnsworth, D. W. (1940). High-speed motion pictures of the human vocal cords. *Bell Laboratories Record*, (18):203–208.
- Fleet, D. and Weiss, Y. (2006). Optical Flow Estimation. In Paragios, N., Chen, Y., and Faugeras, O., editors, *Handbook of Mathematical Models in Computer Vision*, chapter 15, pages 237–257. Springer US.
- Fourcin, A. J. and Abberton, E. (1971). First applications of a new laryngograph. *Medical & biological illustration*, 21(3):172.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202.

- Garcia, M. (1854). Observations on the human voice. *Proceedings of the Royal Society of London*, 7:399–410.
- Gérardin, M. (1994). *Mechanical vibrations : Theory and application to structural dynamics*. Wiley.
- Granados, A., Brunskog, J., Misztal, M. K., Visseq, V., and Erleben, K. (2014). Finite element modeling of the vocal folds with deformable interface tracking. In *Proceedings of Forum Acusticum*.
- Gray, H. (1858). *Anatomy descriptive and surgical*. London: J.W. Parker and Son.
- Gunter, H. E. (2003). A mechanical model of vocal-fold collision with high spatial and temporal resolution. *Journal of the Acoustical Society of America*, 113(2):994–1000.
- Hadwin, P. J., Galindo, G. E., Daun, K. J., Zanartu, M., Erath, B. D., Cataldo, E., and Peterson, S. D. (2016). Non-stationary Bayesian estimation of parameters from a body cover model of the vocal folds. *Journal of the Acoustical Society of America*, 139(5):2683–2696.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
- Hansen, P. C. (1998). *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. Siam.
- Hess, M. M. and Gross, M. (1993). High-speed, light-intensified digital imaging of vocal fold vibrations in high optical resolution via indirect microlaryngoscopy. *Annals of Otology, Rhinology and Laryngology*, 102(7):502–507.
- Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203.
- Hughes, T. J. R. (2000). *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover.
- Ishizaka, K. and Flanagan, J. L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Technical Journal*, 51(6):1233–1268.

- Ishizaka, K. and Matsudaira, K. (1972). *Fluid mechanical considerations of vocal cord vibration*. Monographs 8, Speech Communication Research Laboratory, Santa Barbara, CA.
- Jackson, J. E. (1991). *A User's Guide to Principal Components*. John Wiley & Sons, Hoboken, NJ.
- Jiang, J. J. and Titze, I. R. (1994). Measurement of vocal fold intraglottal pressure and impact stress. *Journal of Voice*, 8(2):132–144.
- Johns, M. M. (2003). Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 11(6):456–61.
- Kaipio, J. and Somersalo, E. (2005). *Statistical and Computational Inverse Problems*. Springer-Verlag New York.
- Kopec, G. (1986). Formant tracking using hidden Markov models and vector quantization. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(4).
- Larsson, H., Hertegård, S., Lindestad, P. a., and Hammarberg, B. (2000). Vocal fold vibrations: high-speed imaging, kymography, and acoustic analysis: a preliminary report. *The Laryngoscope*, 110(12):2117–2122.
- Lautrup, B. (2004). *Physics of Continuous Matter: Exotic and Everyday Phenomena in the Macroscopic World*. CRC Press.
- Lempriere, B. M. (1968). Poisson's Ratio in Orthotropic Materials. *The American Institute of Aeronautics and Astronautics*, 6:2226–2227.
- Lodermeyer, A., Becker, S., Döllinger, M., and Kniesburges, S. (2015). Phase-locked flow field analysis in a synthetic human larynx model. *Experiments in Fluids*, 56(4):1–13.
- Lohscheller, J., Toy, H., Rosanowski, F., Eysholdt, U., and Döllinger, M. (2007). Clinically evaluated procedure for the reconstruction of vocal fold vibrations from endoscopic digital high-speed videos. *Medical Image Analysis*, 11(4):400–413.
- Lucas, B. D. and Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision. *Imaging*, 130:674–679.
- Lucero, J. C., Van Hirtum, A., Rutu, N., Cisonni, J., and Pelorson, X. (2009). Validation of theoretical models of phonation threshold pressure with data from a vocal fold mechanical replica. *The Journal of the Acoustical Society of America*, 125(2):632–5.

- Mittal, R., Erath, B. D., and Plesniak, M. W. (2013). Fluid dynamics of human phonation and speech. *Annual review of fluid mechanics*, 45:437–467.
- Murray, P. R. and Thomson, S. L. (2012). Vibratory responses of synthetic, self-oscillating vocal fold models. *Journal of the Acoustical Society of America*, 132(September):3428–3438.
- Nitsche, J. (1970). Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 36(1):9–15.
- Oertel, M. (1895). Das Laryngo-stroboskop und die laryngo-stroboskopische Untersuchung. *Archiv für Laryngologie und Rhinologie*, 3:1–16.
- Pelorson, X., Hirschberg, A., Van Hassel, R. R., Wijnands, A. P. J., and Augereau, Y. (1994). Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *Journal of the Acoustical Society of America*, 96(6):3416–3431.
- Pinheiro, A. P., Stewart, D. E., MacIel, C. D., Pereira, J. C., and Oliveira, S. (2012). Analysis of nonlinear dynamics of vocal folds using high-speed video observation and biomechanical modeling. *Digital Signal Processing*, 22(2):304–313.
- Proakis, J. G. and Manolakis, D. G. (1996). *Digital Signal Processing (3rd Ed.): Principles, Algorithms, and Applications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Risholm, P., Janoos, F., Norton, I., Golby, A. J., and Wells, W. M. (2013). Bayesian characterization of uncertainty in intra-subject non-rigid registration. *Medical Image Analysis*, 17(5):538–555.
- Rosa, M. D. O., Pereira, J. C., Grellet, M., and Alwan, A. (2003). A contribution to simulating a three-dimensional larynx model using the finite element method. *Journal of the Acoustical Society of America*, 114(5):2893–2905.
- Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinski, B. R., and Afjeh, A. A. (2001). Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *The Journal of the Acoustical Society of America*, 109(4):1616–1630.
- Schön, T. B., Wills, A., and Ninness, B. (2011). System identification of nonlinear state-space models. *Automatica*, 47(1):39–49.

- Schwarz, R., Döllinger, M., Wurzbacher, T., Eysholdt, U., and Lohscheller, J. (2008). Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model. *Journal of the Acoustical Society of America*, 123(5):2717–2732.
- Šidlof, P., Zörner, S., and Hüppe, A. (2015). A hybrid approach to the computational aeroacoustics of human voice production. *Biomechanics and Modeling in Mechanobiology*, 14(3):473–488.
- Simoen, E., Papadimitriou, C., and Lombaert, G. (2013). On prediction error correlation in Bayesian model updating. *Journal of Sound and Vibration*, 332(18):4136–4152.
- Simoncelli, E., Adelson, E., and Heeger, D. (1991). Probability distributions of optical flow. In *Proceedings of Computer Vision and Pattern Recognition*, pages 310–315. IEEE Comput. Sco. Press.
- Simoncelli, E. P. (1993). *Distributed representation and analysis of visual motion*. PhD thesis.
- Spencer, M., Siegmund, T., and Mongeau, L. (2008). Determination of superior surface strains and stresses, and vocal fold contact pressure in a synthetic larynx model using digital image correlation. *Journal of the Acoustical Society of America*, 123(2):1089–1103.
- Story, B. H. and Titze, I. R. (1995). Voice simulation with a body-cover model of the vocal folds. *Journal of the Acoustical Society of America*, 97(2):1249–1260.
- Sutton, M., Mingqi, C., Peters, W., Chao, Y., and McNeill, S. (1986). Application of an optimized digital correlation method to planar deformation analysis. *Image and Vision Computing*, 4(3):143–150.
- Švec, J. G. and Schutte, H. K. (1996). Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*, 10(2):201–205.
- Tao, C., Jiang, J. J., and Zhang, Y. (2006). Simulation of vocal fold impact pressures with a self-oscillating finite-element model. *Journal of the Acoustical Society of America*, 119(6):3987–3994.
- Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. *Journal of the Acoustical Society of America*, 83(4):1536–1552.
- Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America*, 85(2):901–906.

- Titze, I. R. (2006). *The Myoelastic Aerodynamic Theory of Phonation*. National Center for Voice and Speech.
- Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory. *Journal of the Acoustical Society of America*, 123(5):2733–2749.
- Titze, I. R. and Hunter, E. J. (2007). A two-dimensional biomechanical model of vocal fold posturing. *Journal of the Acoustical Society of America*, 121(4):2254.
- Titze, I. R., Svec, J. G., and Popolo, P. S. (2003). Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues. *Journal of Speech Language and Hearing Research*, 46(4):919–932.
- Titze, I. R. and Talkin, D. T. (1979). A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. *Journal of the Acoustical Society of America*, 66(1):60–74.
- Van Den Berg, J., Zantema, J. T., and Doornenbal, R. J. (1957). On the Air Resistance and the Bernoulli Effect of the Human Larynx. *Journal of the Acoustical Society of America*, 29(5):626–631.
- Van Hirtum, A., Grandchamp, X., and Pelorson, X. (2009). Moderate Reynolds number axisymmetric jet development downstream an extended conical diffuser: Influence of extension length. *European Journal of Mechanics, B/Fluids*, 28(6):753–760.
- Weiss, Y. and Fleet, D. J. (2002). Velocity likelihoods in biological and machine vision. In *Probabilistic Models of the Brain: Perception and Neural Function*, pages 81–100. MIT Press.
- Wittenberg, T., Moser, M., Tigges, M., and Eysholdt, U. (1995). Recording, processing, and analysis of digital high-speed sequences in glottography. *Machine Vision and Applications*, 8(6):399–404.
- Wriggers, P. (2002). *Computational contact mechanics*. John Wiley & Sons.
- Xue, Q., Mittal, R., Zheng, X., and Bielamowicz, S. (2012). Computational modeling of phonatory dynamics in a tubular three-dimensional model of the human larynx. *Journal of the Acoustical Society of America*, 132(3):1602–1613.
- Yang, A., Lohscheller, J., Berry, D. A., Becker, S., Eysholdt, U., Voigt, D., and Döllinger, M. (2010). Biomechanical modeling of the three-dimensional aspects of human vocal fold dynamics. *Journal of the Acoustical Society of America*, 127(2):1014–1031.

- Yang, A., Stingl, M., Berry, D. A., Lohscheller, J., Voigt, D., Eysholdt, U., and Döllinger, M. (2011). Computation of physiological human vocal fold parameters by mathematical optimization of a biomechanical model. *Journal of the Acoustical Society of America*, 130(2):948–64.
- Zañartu, M., Erath, B. D., Peterson, S. D., Hillman, R. E., and Wodicka, G. R. (2013). Modeling incomplete glottal closure due to a posterior glottal opening and its effects on the dynamics of the vocal folds. *The Journal of the Acoustical Society of America*, 133(5).
- Zhang, K., Siegmund, T., and Chan, R. W. (2006). A constitutive model of the human vocal fold cover for fundamental frequency regulation. *Journal of the Acoustical Society of America*, 119(2):1050–1062.
- Zheng, X., Xue, Q., Mittal, R., and Beilamowicz, S. (2010). A coupled sharp-interface immersed boundary-finite-element method for flow-structure interaction with application to human phonation. *Journal of Biomechanical Engineering*, 132(11):111003 (1–12).

www.elektro.dtu.dk

Department of Electrical Engineering

Acoustic Technology

Technical University of Denmark

Ørsted's Plads

Building 348

DK-2800 Kgs. Lyngby

Denmark

Tel: (+45) 45 25 38 00

Fax: (+45) 45 93 16 34

Email: info@elektro.dtu.dk